

DTIC FILE COPY

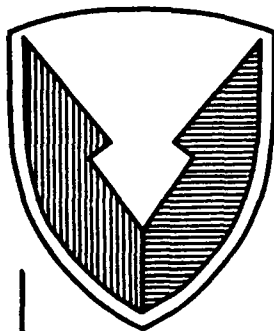
4

AD-A210 974

RD & E

C E N T E R

Technical Report



No. 13432

ROBOT VEHICLE VIDEO IMAGE COMPRESSION

PHASE I SBIR FINAL REPORT

CONTRACT NUMBER DAAE07-88-C-R068

JANUARY 1989

DTIC
SELECTE
AUG 08 1989
S D

Carl F. R. Weiman
Transitions Research Corporation
15 Great Pasture Road
Danbury, CT 06810

By

APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION IS UNLIMITED

U.S. ARMY TANK-AUTOMOTIVE COMMAND
RESEARCH, DEVELOPMENT & ENGINEERING CENTER
Warren, Michigan 48397-5000

89

8

08

087

NOTICES

This report is not to be construed as an official Department of the Army position.

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Mention of any trade names or manufacturers in this report shall not be construed as an official endorsement or approval of such products or companies by the U.S. Government.



Destroy this report when it is no longer needed. Do not return it to the originator.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release: Distribution is Unlimited		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			4. PERFORMING ORGANIZATION REPORT NUMBER(S)		
5. MONITORING ORGANIZATION REPORT NUMBER(S)			6a. NAME OF PERFORMING ORGANIZATION Transitions Research Corp.		
6b. OFFICE SYMBOL (If applicable)			7a. NAME OF MONITORING ORGANIZATION U.S. Army Tank-Automotive Command		
7b. ADDRESS (City, State, and ZIP Code) 15 Great Pasture Road Danbury, CT 06810			7c. ADDRESS (City, State, and ZIP Code) R, D&E Center Warren, MI 48387-5000		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION			8b. OFFICE SYMBOL (If applicable)		
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER			10. SOURCE OF FUNDING NUMBERS		
10. SOURCE OF FUNDING NUMBERS			PROGRAM ELEMENT NO.		
10. SOURCE OF FUNDING NUMBERS			PROJECT NO.		
10. SOURCE OF FUNDING NUMBERS			TASK NO.		
10. SOURCE OF FUNDING NUMBERS			WORK UNIT ACCESSION NO.		
11. TITLE (Include Security Classification) Robotic Vehicle Video Image Compression					
12. PERSONAL AUTHOR(S) Carl F. R. Weiman (Transitions Research Corporation)					
13a. TYPE OF REPORT Final		13b. TIME COVERED FROM 88 JUL TO 89 JAN		14. DATE OF REPORT (Year, Month, Day) 89 JAN	
15. PAGE COUNT 66					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD			GROUP		
SUB-GROUP			robotics, data compression, video compression		
			teleoperation, remote vehicle, communications		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This Phase I SBIR study analyzed and simulated in software a new approach to image data compression for remote vehicle driving. Results show feasibility for 1,000-to-1 compression rates for full color, wide field of view, real time imagery. The approach is based on a three stage process. The first stage consists of reducing resolution of peripheral imagery to match the drop in human visual resolution away from the center of view. This is accomplished by resampling video camera output in real time. The second stage consists of coding color and contrast edges in separate channels. The third stage consists of applying a simple, efficient data compression technique. A hybrid DCT/DPCM data compression method proved to be the most suitable. All three stages of the compression process and reconstruction of imagery can be implemented in real time on equipment which can be compactly mounted in remote vehicles. This is a fundamentally new and valuable attack on the problem of remote driving, which has been with us for decades. The solution is critical for deployment of Army remote systems.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> OTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Brian M. Novak			22b. TELEPHONE (Include Area Code) (313) 574-8678		22c. OFFICE SYMBOL AMSTA-RRT

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
EXECUTIVE SUMMARY	4
1.0 INTRODUCTION	5
2.0 OBJECTIVE	7
3.0 CONCLUSIONS	7
4.0 RECOMMENDATIONS	8
4.1 Field System Specification	9
4.2 Phase II SBIR Project	20
5.0 DISCUSSION	24
5.1 Human Visual Factors	25
5.2 Logarithmic Mapping and Reconstruction	38
5.3 Edge and Color Coding and Reconstruction	48
5.4 Image Compression Techniques	55
5.5 Summary	58
 Bibliography	 59
Photos	61

LIST OF FIGURES AND PHOTOS

	Page
Figure 4-1 Camera Fields of View (top view)	11
Figure 4-2 Operator Panoramic Display	11
Figure 4-3 Data Flow of Recommended Field System	13
Figure 4-4 Resampling Geometry on Display Screens	14
Figure 4-5 Real Time Laboratory Prototype System	22
Figure 5-1 Close-up of Text Near the Fovea	26
Figure 5-2 Text Showing Decreasing Resolution Peripherally	27
Figure 5-3 Text With Moved Center of Fixation	28
Figure 5-4 Spatio-Temporal Resolution Envelope for Humans	31
Figure 5-5 Human Receptive Field Diameters	31
Figure 5-6 Human Receptive Field Distribution	32
Figure 5-7 Color Perception Parameter Spaces	34
Figure 5-8 Nominal Viewing Distance for Video Display	36
Figure 5-9 Video Display Screen within Human Field of View	37
Figure 5-10 Conformal Mapping for Resampling Video Imagery	39
Figure 5-11 Geometry of Image Reconstruction Interpolation	42
Figure 5-12 Geometry of Gaussian Interpolation	43
Figure 5-13 Separation of 2-D Gaussian into 1-D Gaussians	45
Figure 5-14 Integral of 1-D Gaussian	46
Figure 5-15 Color Discrimination on CIE Diagram	49
Figure 5-16 Edge Detection Filter Template	51
Figure 5-17 Grossberg's Directional Edge Detector	52
Figure 5-18 Sharpening of Edges from the Log Domain	54
Figure 5-19 16-by-16 Pixel Domain of Hybrid Data Compression	56
Photo 1 - Raw Video Color Image of Typical Driving Scene	61
Photo 2 - Reconstructed Low Resolution Color Image	61
Photo 3 - Reconstructed Edge Density Image	62
Photo 4 - Composite Color and Edge Image	62
Photo 5 - Composite Image with Different Center of Attention	63
Photo 6 - Log-Mapped Image in Upper Left of Picture	63
Photo 7 - Reconstruction with Zeroth Order Interpolation	64
Photo 8 - Reconstructed High Resolution Color Image	64
Photo 9 - Reconstructed Edges Smeared into Blobs	65
Photo 10 - Gradient Image of the Smeared Edges	65
Photo 11 - Composite Image using Linear Feature Data	66

ROBOT VEHICLE VIDEO IMAGE COMPRESSION

EXECUTIVE SUMMARY

The Army is developing robotic vehicles for a variety of missions in hostile or hazardous environments. Video transmission to command stations is a key element for observation and vehicle control. High bandwidth of conventional video requires line-of-sight or optic fiber communications, which reduce field effectiveness. Video data compression is desirable.

This Phase I SBIR study analyzed and simulated in software a new approach to image data compression for remote vehicle driving. Results show feasibility for 1,000-to-1 compression rates for full color, wide field of view, real time imagery.

The approach is based on a three stage process.

The first stage consists of reducing resolution of peripheral imagery to match the drop in human visual resolution away from the center of view. This is accomplished by resampling video camera output in real time.

The second stage consists of extracting image cues which are key to remote driving, and coding them to match characteristics of human vision. Color and contrast edges are coded in separate channels.

The third stage consists of applying a simple, efficient data compression technique. A hybrid discrete cosine transform - differential pulse code modulation (DCT/DPCM) data compression method proved to be the most suitable.

All three stages of the compression process and reconstruction of imagery can be implemented in real time on equipment which can be compactly mounted in remote vehicles. A few key components have already been implemented in hardware on printed circuit boards using off-the-shelf components.

This is a fundamentally new and valuable attack on the problem of remote driving, which has been with us for decades. The solution will be critical for effective deployment of remote systems which the Army is developing.

1.0 INTRODUCTION

There is a growing recognition of the need for autonomous and remotely controlled vehicle systems for reconnaissance, force multiplication in tactical engagement, mine detection and clearing, and other applications. Current Robotics programs include the Army's Robotic Command Center (RCC), Technology Enhancements for Autonomous Machines (TEAM) program, Wiesel program, Robotic Obstacle Breaching Assault Tank (ROBAT), MIRADORS, and the Marine Corps Teleoperated Vehicle (TOV) program.

The ALV and other autonomous vehicle research programs have shown that full autonomy is not practical in the near term. Remote teleoperation is necessary. For effective remote control in natural terrain, video imagery is a key element of operator feedback. No other sensor can provide better data to the operator for rapid execution of driving and other mission supporting tasks such as obstacle avoidance, threat and target detection, reconnaissance, and path planning.

The problem with video transmission is that it requires such high bandwidths that optic fiber or line-of-sight microwave communication channels are necessary. The former restricts range and maneuverability, the latter puts vehicles in high exposure positions as well as restricting range.

Data compression could solve the high bandwidth problem. A number of attempts have been made, based on slow transmission or data compression algorithms. The former eliminates motion cues and introduces delays which degrade mobility. The latter do not provide enough reduction to eliminate line-of-sight constraints.

TRC has developed a new technique which this study indicates can provide 1,000-to-1 compression rates for full color, wide field of view, real time imagery.

The approach is based on a three stage process.

The first stage consists of reducing resolution of peripheral imagery to match the drop in human visual resolution away from the center of view. This is accomplished by resampling video camera output in real time, using a high resolution center whose position on the screen is controlled by the operator.

The second stage consists of extracting image cues which are key to remote driving, and coding them to match characteristics of human vision. Color and contrast edges are coded in separate channels.

The third stage consists of applying a simple, efficient data compression technique. A hybrid discrete cosine transform - differential pulse code modulation (DCT/DPCM) data compression method proved to be the most suitable.

All three stages of the compression process and reconstruction of imagery can be implemented in real time on equipment which can be compactly mounted in remote vehicles. A few key components have already been implemented in hardware on printed circuit boards using off-the-shelf components.

The scope of this study was a 6 month Phase I SBIR contract which included analysis, design and software simulation on TACOM provided remote driving imagery. The following pages describe the results.

TRC was founded in 1984 to provide consulting and engineering development services in new applications of robots and automation in the service sector of the economy. Examples of projects carried out by TRC are the development of a navigation and control system for an autonomous robot for floor cleaning, development of a robot for materials transport in hospitals and nursing homes, and robotics and vision systems for the NASA Space Station.

2.0 OBJECTIVE

The subject of this study was to analyze and determine the feasibility of an image compression system for visual operation of remote vehicles. The goal was to reduce bandwidth sufficiently to eliminate line-of-sight restrictions on RF (radio frequency) transmission.

The design concept was to be based on a three stage process. The first stage consists of reducing resolution of peripheral imagery to match the drop in human visual resolution away from the center of view. The second stage consists of extracting image cues which are key to remote driving, and coding them to match characteristics of human vision. The third stage consists of applying a simple, efficient data compression technique.

The objectives of the study included defining specific algorithms for processing video data through the three compression stages, simulating these in software, and determining their net compression rates and suitability for remote driving.

3.0 CONCLUSIONS

The study successfully concluded that image data compressions exceeding 1,000-to-1 could be achieved by implementing the three stage process of reducing peripheral resolution, coding perceptual channels, and applying a standard data compression technique to the result. Software simulation of TACOM provided remote driving imagery verified key elements.

Analysis of remote driving video tapes and RV (remote vehicle) camera specifications showed that existing camera and display configurations for remote vehicle driving are grossly mismatched to the parameters of operator perception. The total field of view is too narrow. Even within the limited field of view, data is transmitted that is not perceivable to the operator. The result is wasted bandwidth and inadequate visual cues for path planning, threat detection, and obstacle avoidance.

Adequate visual context can be provided by increasing the total field of view to 60° by 120°. Real time (30 frames per second) imagery is necessary to engage the operator's depth perception cues from image motion. Color is vital in providing environmental context and figure-background discrimination. Brightness contrast edges are critical for perceptual resolution.

The first stage of compression can be accomplished by resampling the video signal at decreasing peripheral resolution to match the human resolution perception gradient. The key to usefulness is to allow the operator to instantly move the center of resolution to any point of interest within the field of view. The conclusions from analysis and simulation showed that 25-to-1 reduction in pixel count is possible.

It was concluded that the second stage of compression, separate coding of color and contrast channels, would yield an additional 8-to-1 compression ratio. Color is perceived in human vision at only one quarter of the resolution of contrast edges, and may therefore be encoded with fewer data elements. Contrast edges can be represented as binary data, requiring less data per element than color.

For the third stage of data compression, a hybrid DCT/DPCM (discrete cosine transform / differential pulse code modulation) scheme best satisfied requirements for compression ratio, efficiency, and robustness. Hybrid data compression yields an additional 8-to-1 compression ratio.

The results show that full color, wide field of view, 30 frame per second, high central resolution imagery can be provided for successful remote vehicle driving at data transmission rates compressed by better than 1,000-to-1 compared to standard color video transmission. Data rates under 100 kilobits/second are reasonable for this system.

A compact mobile field system meeting the above requirements could be fabricated using off-the-shelf integrated circuit components.

4.0 RECOMMENDATIONS

We recommend a development project under a Phase II SBIR for experimentally determining optimal parameters, algorithms, and hardware design for a proposed field system. Section 4.2 describes the development project.

We recommend design of a field system with a field of view 120° wide and 60° high. To get the required high resolution at the center and wide field of view, three color cameras with abutting fields are recommended. This is compatible with the present RV configuration. Three corresponding panoramic color displays should be provided for the operator.

Section 4.1 below describes the recommended field system. There are three major components of data compression.

First, reduced resolution peripherally matches the perceptual resolution gradient of the human operator, compressing pixel count by 25-to-1.

Second, image information should be separated into data channels whose features match those of the human visual channels. Color data can be transmitted at low spatial resolution and contrast edges can be transmitted at higher spatial resolution. These perceptual channel codings yield an additional compression factor of 8-to-1.

Third, the results of the two preceding steps should be input to a standard data compression algorithm chosen for robustness and simplicity of implementation. This yields an additional 8-to-1 compression ratio.

The net result of these steps is a system which can transmit wide angle, high central resolution color video images at 30 frames per second over channels running at somewhat less than 100 kilobits per second.

4.1 Field System Specification

Observations of TACOM supplied remote driving tapes and analysis of human vision suggest that much wider fields of view than provided by a single conventional camera could greatly improve operator perception of the 3-D environment using peripheral vision and optic flow. A short vertical field of view prevents the simultaneous visual reference of horizon and immediate path. A narrow field of view deprives the operator of the following mission critical inputs:

- a panorama of the horizon for orientation
- a view of the road ahead when it curves
- a view of immediate alternative paths of opportunity when an obstacle lies directly ahead
- peripheral warning of threats
- medium to long range terrain assessment for path planning

Wide field of view lenses suffer from severe vignetting (decrease of light-gathering power peripherally) and radial (fisheye) distortion. Although these could be partially overcome by lookup tables, the fixed resolution of CCD (or any other) imaging chips inherently yields reduced data resolution per solid angle as the field of view is increased. A simple multiple camera system compatible with the configuration of the TACOM RV system provides a solution.

4.1.1 Camera and Display Configuration

Figure 4-1 illustrates the geometry of a field of view 120 degrees wide and 60 degrees high, provided by three standard CCD camera with 8 mm lenses. The fields of view are abutted to provide the operator display panorama illustrated in **figure 4-2**. The cameras are rotated 90 degrees so that scan lines are vertical. This technique is used in some computer displays and video games where large vertical fields of view are desired.

A field of view 120 degrees wide covers a significant portion of human peripheral vision. A 60° vertical field of view provides frontal coverage from 5 feet in front of the vehicle on the ground to above the horizon. The near range coverage is based on cameras mounted 3 feet above the ground, pointed at the horizon. Viewfield width on the ground is 17 feet, providing excellent local obstacle and terrain perception.

Three displays are arranged to subtend to the operator the same visual angle as viewed by the cameras. Thus 17 inch displays would be viewed from 12 inches. These ranges are much too close for effectively viewing standard monitors. However, optical elements could provide the same field of view without close focus. Visual flight simulators use "infinity optics" in the form of concave mirrors to achieve a wide field of view without close focus. A planar holographic lens, or helmet mounted displays could also be used.

As an alternative to current video display technology, HDTV (high definition television) monitors could serve as superior displays for remote driving. This emerging technology proposes to replace the current 525 line U. S. television broadcast standard with 1,050 to 1,250 line video imagery [Business Week, Jan 30, '89]. Although not presently available commercially, prototypes are working in Japan and the United States. The prospect of enormous commercial markets is spurring rapid development which may well yield components applicable to fieldable remote driving systems within the next two years.

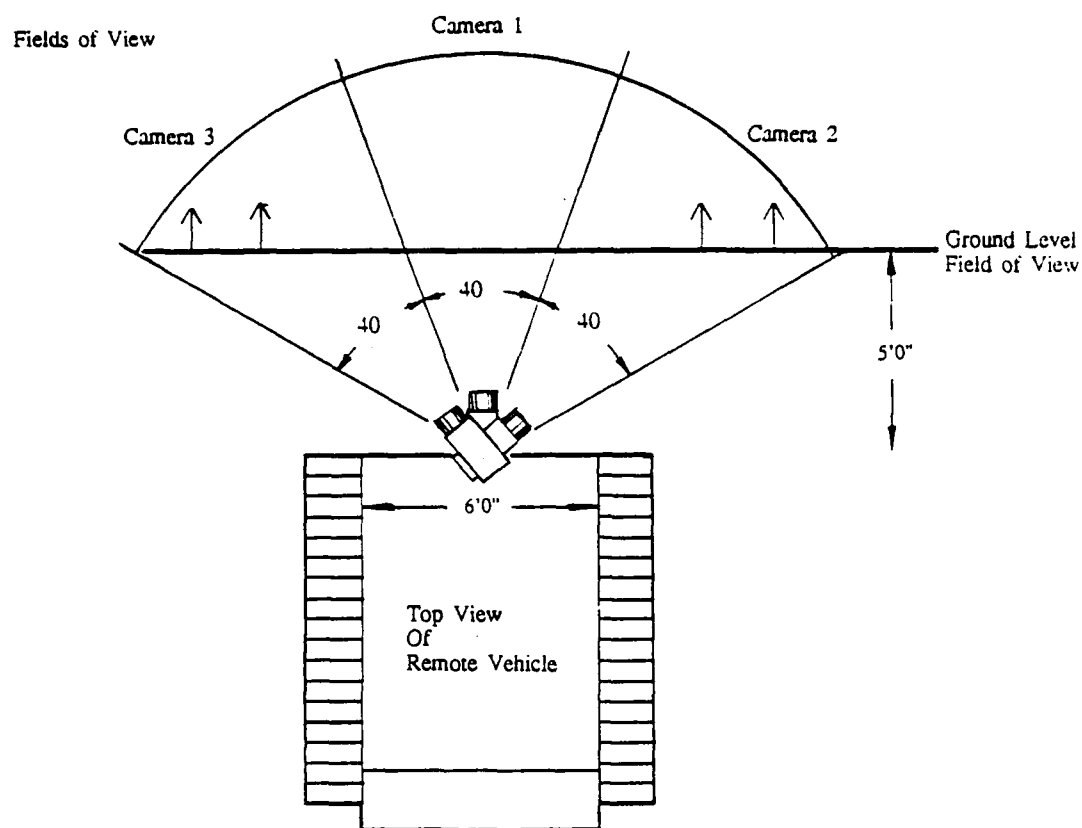


Figure 4-1 Camera Fields of View (top view)

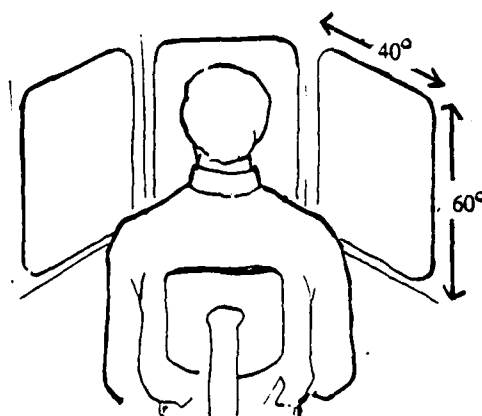


Figure 4-2 Operator Panoramic Display

4.1.2 Processing Algorithms

The recommended processing sequence upstream of transmission consists of video image remapping to reduce pixel count, coding of color and contrast edges, followed by data compression. **Figure 4-3** illustrates the data flow.

In the first stage of compression, video data is resampled via the logarithmic mapping tables to reduce pixel count. These effect a resampling geometry which is a rotationally symmetric pattern whose cells increase in size linearly with distance from the center of the image. The result is known as conformal logarithmic or log polar mapping. The mapping sampling pattern is spread panoramically across the view fields of all three cameras, as illustrated in **figure 4-4**. The center is chosen by the operator, as described in section 4.1.3 below.

The first stage splits into two parallel paths, one for black and white, and the other for color at lower resolution. Black-and-white imagery is mapped using a pattern with 128 radial wedges of pixels. The inner ring of the 128 wedge pattern is forty video lines in diameter for a 480-by-512 pixel video sensor array. Within this ring, original video is preserved. This is roughly double the diameter of the human fovea. Each video line subtends less than 6 arcminutes, the diameter of human receptive fields in the fovea, the central region of uniform high resolution in human vision. Human receptive fields double in size for each doubling in radius outside the fovea. The growth in size of resampling cells in the 128 wedge pattern in **figure 4-4** matches this human resolution gradient.

The color image is mapped using a pattern with 32 radial wedges of pixels. This represents a 4-by-4 subsampling of black and white contrast data, which corresponds closely to the ratio of color vision resolution to brightness contrast resolution in human vision, as described in section 5.1.3.

The second stage of compression for black and white imagery consists of applying an edge detection filter to log mapped brightness data in the upper path of **figure 4-3**. The result is a map of local high contrast intensities. Color processing consists of averaging all color values within a resampling cell.

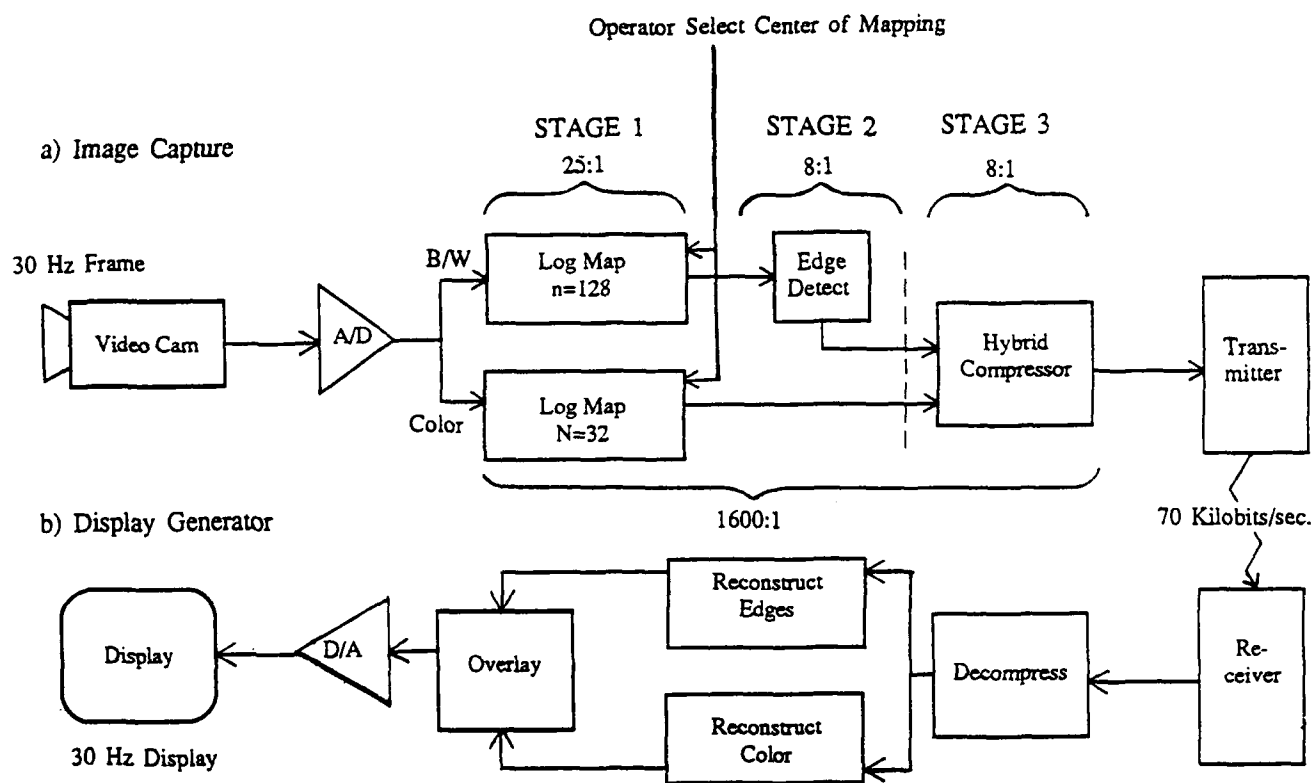


Figure 4-3. Data Flow of Recommended Field System

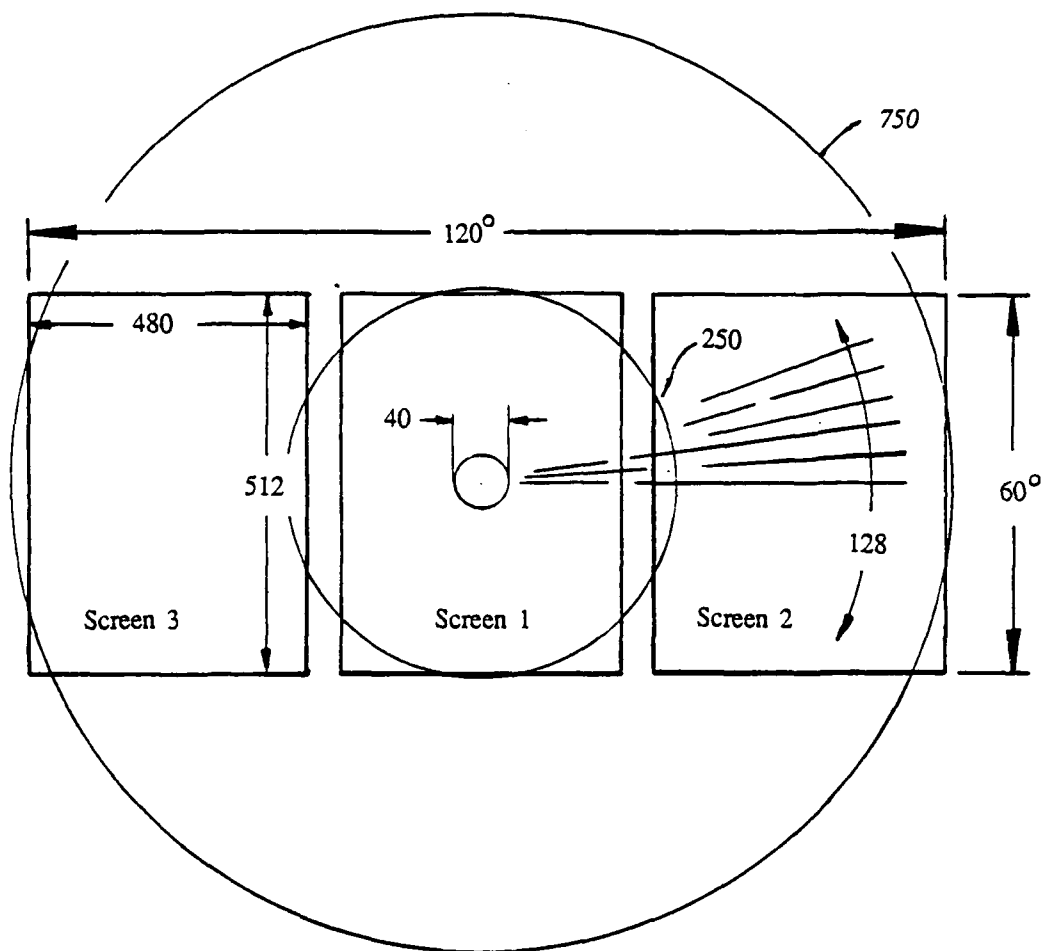


Figure 4-4. Resampling Geometry Superimposed on Display Screens

The third stage of compression is a hybrid data compression technique as follows. Edge contrast and color data are blocked into 16-by-16 element data arrays. Each row of 16 elements is discrete cosine transformed, and the resulting coefficients are DPCM (differential pulse code modulation) encoded. The output digital data stream is transmitted from the RV.

The signal transmitted by the RV is received at a remote command station, and reproduced as a digital data stream. The DPCM and Discrete Cosine Transformation steps are inverted, yielding the logarithmic mapped image. Color data is exponentially mapped into a display buffer using Gaussian interpolation, described in section 5.2.2, or bicubic interpolation. Edge contrast data is reconstructed also by inverse logarithmic mapping, and sharpened using image processing such as thinning and thresholding.

Edge data is superimposed on color data by overlaying black edge data on the color data, yielding a cartoon-like quality image. More subtle overlays could be achieved by inverting or brightening underlying color imagery where contrast edges occur.

Photo 1 illustrates a raw video color image of a typical driving scene. **Photo 2** illustrates the reconstructed low resolution color image. **Photo 3** illustrates the reconstructed edge density image. **Photo 4** illustrates the processed composite edge and color image as seen by the operator at a remote control station. The reader should view **photo 4** from no more than 6 inches to match scene detail with perceptual resolution.

4.1.3 Operator Interface

The perceptual match between displayed data and the distribution of human visual resolution is valid only if the operator is looking directly at the center of the mapping. Off center viewing of **photos 2, 3, or 4** reveals a grossly blurred image. The restriction of a fixed viewing center is totally unacceptable. Indeed, we constantly move our eyes around in the visual field to bring points of interest into high resolution. Therefore, the ability to move the center of the mapping around the image to focus on points of interest is critical to the utility of the proposed system. **Photo 5** illustrates the reconstruction of the same image as **photo 1**, but with the fovea centered on the small building at the upper right.

In the teleoperator environment, there is information flowing in both directions; visual data from the vehicle and control data from the operator. Such a system could easily accommodate commands from the operator to change the center of mapping in the vehicle video processor to coincide with the operator's point of gaze on the display screen. The most transparent implementation would be to use an eye tracker in the operator station to measure the direction of gaze of the operator. Screen coordinates (9 bits each for x and y) of the gaze direction would then be computed and transmitted to the vehicle. At the vehicle, this offset would be added to the nominal center of screen coordinates to serve as the new center of the mapping. This operation is trivial and can be carried out at frame rates without changing lookup tables. The operator would perceive the scene at full resolution at all times, wherever he was looking.

Although a mouse or joystick would provide adequate control of focus of attention, the operator's hands and feet may be occupied with driving the remote vehicle. We recommend that the operator move his focus of attention under the control of a head tracker. A head tracker is mechanically simpler and easy to implement than an eye tracker. A simple design could be based on a helmet mounted LED (light emitting diode) aiming at a wide field of view black-and-white CCD video camera.

We also recommend that the operator have control over data transmission rates and image quality. Higher quality images at slower update rates could be achieved by transmitting 8-bit per pixel brightness data rather than edge data. Update rate would slow from 30 frames per second to 5 frames per second.

Another possibility would be to have a higher (say 1 megabit per second) transmission bandwidth available for the operator to call on if needed, preempting other sensor communications from the remote vehicle. Update rate would remain at 30 frames per second and bandwidth would vary. If other data transmission requirements were compatible with this occasional priority use for video, this might be preferable to reduced frame rates.

4.1.4 System Summary and Data Rates

Figure 4-4 depicts three camera input screens of video data, overlaid with appropriate resampling ring radii. Input video data is 480-by-512, or 245,760 pixels. The fovea (uniform central disk of the mapping) is shown centered on screen 1, with the understanding that it may be moved freely among the three screens. Wherever it moves, mapping domains in the panoramically adjacent screens are concentric with the fovea and consistent in sampling rate.

The sampling rate for high resolution edge data is $n = 128$ elements in each ring, starting at 20 video lines from the center. The fovea, whose pixel domains are unaltered, thus contains

$$F1 = \pi \cdot r \cdot r = 400\pi = 1256 \text{ pixels.}$$

The number of resampling rings outside the fovea within screen 1 is

$$m = (n/2\pi) \cdot \ln(r_{\max}/r_{\min}) = 51 \text{ rings}$$

where $n = 128$ resampling wedges, $r_{\min} = 20$, and $r_{\max} = 250$. The equations for ring number and pixel count are found in [Weiman 88a].

The number of peripheral resampling cells in screen 1 is thus

$$P1 = m \cdot n = 51 \cdot 128 = 6528.$$

The total number of resampling cells in screen 1 is thus

$$N1 = F1 + P1 = 1256 + 6528 = 7784 \text{ pixels.}$$

This represents a 32-to-1 pixel count reduction for screen 1.

The number of pixels in screens 2 and 3 can be calculated by computing the average density of resampling cells per video line within their boundaries, and multiplying by the areas of the screens, as follows.

Conservatively estimate the inner and outer boundaries of screen 2 to be at 250 and 750 video lines from the fovea.

The resampling ring numbers associated with these distances are

$$m = 51 \text{ (as before)}$$

and

$$m + dm = 73$$

where

$$dm = (n/2\pi) \ln(750/250) = 22.$$

The mean ring number is

$$q = (51+73)/2 = 62 .$$

The distance from the fovea center at this ring number is

$$r = 20 \exp(2\pi q/n) = 419 .$$

The diameter of a resampling cell at this radius is

$$d = r \cdot 2\pi/n = 21 \text{ video lines.}$$

Thus average resampling cell area is

$$d \cdot d = 441 \text{ pixels.}$$

Now, the area of screen 2 is

$$A = 512 \cdot 480 = 245,760 \text{ pixels,}$$

so the number of resampling cells per screen is

$$N_2 = N_3 = A/(d \cdot d) = 557 .$$

Adding these to the count of screen 1 cells yields a total of

$$N = N_1 + N_2 + N_3 = 7,784 + 557 + 557 = 8,898$$

resampling cells, a 14% increase over screen 1 resampling cells alone. Net compression ratio in pixel count for all three screens over screen 1 video alone is about 28 to 1.

Thus, we can conservatively claim 25-to-1 compression ratio for this step and leave the door open for slightly higher resampling rates.

Edge detection reduces each resampling cell to a single bit. Color at 16 bits per sampling cell cancels out against the 16:1 size differential of color versus edge spatial resolution.

Thus 16 bits for each of 8,898 pixels reduces to 2 bits each, an 8-to-1 compression ratio, yielding

$$B = 8,898 * 2 = 17,796$$

bits per frame or

$$B \times 30 = 533,880$$

bits per second for screens 1, 2, and 3 outputting resampled color and edge data at 30 frames per second.

The hybrid data compression algorithm described in section 5.4.1 yields another factor of 8 compression so that

$$R = 533,880 / 8 = 66,735$$

bits per second is the net communication rate. This leaves room for error correction coding redundancy or encryption overhead comfortably within 100 kilobits per second.

The net compression ratio from logarithmic resampling, edge and color channel coding, and hybrid data compression is

$$C = 25 * 8 * 8 = 1600:1$$

compared against one screen of 512x480 pixels of 16 bit color imagery.

The data transmission rate could be reduced further to 17,800 kilobits/second by updating imagery at only 8 frames per second. Alternatively, at 4 frames per second, full 16 bit color at 128 elements per ring resolution could be transmitted.

4.2 Phase II SBIR Project

The successful results of Phase I analysis and simulation suggest development of a realtime video image compression system with better than 1000-to-1 compression rates. The next logical step is to build a realtime laboratory prototype with sufficient flexibility to experiment with parameters and alternative algorithms in order to choose the best candidates for a field deployable system. Phase II should consist of construction and test of the lab system, and specification of the deployable system.

4.2.1 Realtime Prototype System

The overall goal of Phase II is to design and build a real-time (30 frame per second) prototype video compression/reconstruction system for delivery to TACOM. This system will consist of image capture, processing, compression, and reconstruction components in a laboratory resident configuration. Interfaces for video communication with field equipment are available to enable real-time remote driving experiments by TACOM using the prototype equipment.

As a prototype, this system will be larger and more flexible than the field deployable system to be built in Phase III. The purpose of the prototype is to tune critical image processing parameters and experiment with alternative techniques. In the experimental scenario, technicians could switch between different compression algorithms, or tune parameters to evaluate performance tradeoffs. Evaluation experiments range from subjective evaluation of output generated from videotape input, to actual remote driving experiments. The Phase III system will incorporate the best algorithms and optimal parameters in a compact, portable, mil-spec configuration.

The host for the Phase II prototype system will be a low cost real-time commercial image processing system such as a DataCube, with appropriate modular processing boards. TRC will design and fabricate prototype boards as needed for performing log mapping and reconstruction of imagery.

Figure 4-5 is a system diagram of the laboratory prototype. Video from the remote vehicle or other source is input to the host image processing system which contains A/D converters and frame buffers. The video input signal should be equivalent to full NTSC color. Its source could be CCD color camera, videotape, or digital color imagery.

A maximum of three visually adjacent video channels would provide the full field of view commensurate with the field deployable system. A single channel is recommended for a baseline prototype system. The extra complexity of processing equipment and engineering design for three channels will be costed as an option in the Phase II proposal.

In parallel with the video input is data specifying the position of the focus of attention of the operator. This comes from a trackball, mouse, joystick, head-tracker, or eye-tracker. The input video image is logarithmically mapped, with the center of the mapping at the specified focus of attention, which can be updated every 33 milliseconds as the operator moves his focus of attention. In figure 4-5 the mapper is shown lateral to the image processing system to indicate that remapping is efficiently executed using digital communication channels intrinsic in the image processing system. Mapped resolution, which is the density at which the original image is logarithmically resampled, depends on the lookup table load in this mapper. The table can be generated using software in the image processing system host. Thus mapped resolution can be flexibly assigned for experimental evaluation.

Color signals are mapped at low resolution and put directly into an output buffer, requiring none of the processing power of the image processing system. In parallel, black and white imagery is digitized at higher resolution, and processed through image processing system hardware to detect or enhance images, and threshold. Color and edge data is then output to the data compressor, which would implement the discrete cosine transform and DPCM coding, or other standard method, in real time.

The output of the data compressor is input to a communications buffer which outputs a high speed serial TTL level digital signal (10 kilobaud to 1 megabaud). This data is suitable for input to an RF transmitter, such as the IMSCO ADC-MX transmitter. The transmitter is not part of the prototype system, but can be provided by the Army for experiments.

The components described above correspond to the RV (remote vehicle) resident components of a field system. The RCC (command center) components consist of a receiver, such as the IMSCO ADC-MR, and downstream image reconstruction equipment. In the prototype, the receiver (not to be provided under the phase II contract) would re-create the high speed digital data signal which was input to the transmitter. The data expander then generates log transformed digital data and dumps it into frame buffer memory.

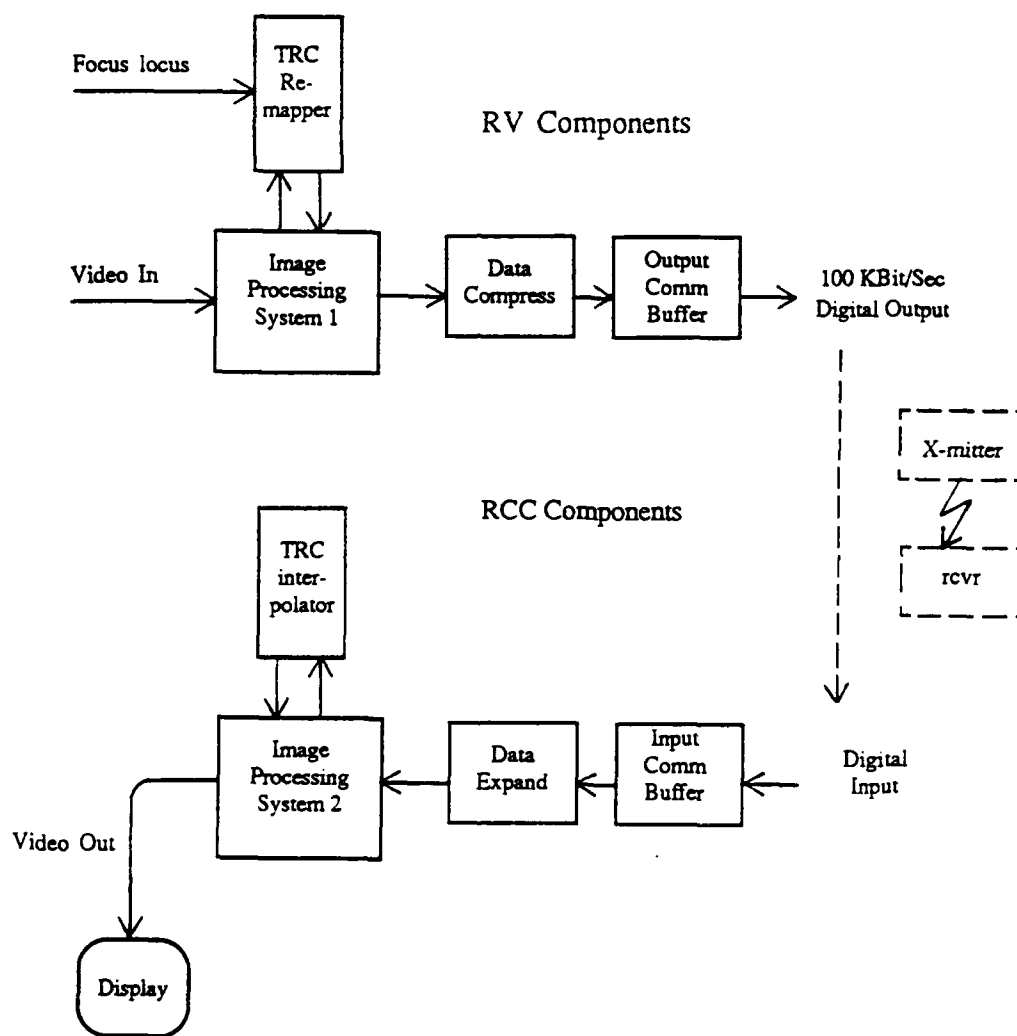


Figure 4-5. Real Time Laboratory Prototype System

The TRC interpolator performs inverse logarithmic mapping (for example, using Gaussian interpolation) and generates black and white edge data and color data, which are merged by logic in a second image processing system.

For remote driving experiments during Phase II, the Army would provide a conventional real-time video input link to a remote vehicle. This could be an optic fiber cable, or line-of-sight microwave signal. It is understood that in the ultimate field system, the destination of such a signal would be a processor within the remote vehicle, and that compressed imagery would be transmitted to the operator at much lower frequencies.

In laboratory experiments, video input from tape or camera is presented to the video input of the RV-side components. Processed output from the communications buffer is coupled directly to the digital input communications buffer on the RCC-side components, as illustrated by the vertical dashed line in figure 4-5. Reconstructed color imagery is displayed to the operator.

In Phase II field experiments, a fiber optic or line-of-sight microwave video receiver inputs video to the RV-side components. The source of the imagery is a live camera on a remote vehicle. Again, the digital output of the RV-side components serves as input to the RCC side components. If the RV is controllable, the experimenter operates the remote controls, which generates control signals transmitted to the vehicle via RV control channels.

4.2.2 Specification of Phase III Field System for RCC-RV

Experiments with the prototype system will determine preferred visual resolution parameters, edge detection algorithms, thresholds, and reconstruction algorithms. These will drive requirements for the field system design, which will be defined towards the end of Phase II.

Aside from viable realtime imagery, essential requirements for the field system are mil-spec hardware, compactness, reliability, and graceful degradation of reconstructed images in the presence of transmission channel noise.

From an operational viewpoint, the field system should permit flexible selection of image quality and focus of attention, depending on local visual and transmission environmental conditions. That is, the operator should be able to trade-off color, gray-scale range, resolution, and refresh rate against each other depending on local conditions.

5.0 DISCUSSION

The key to image compression in this project is to efficiently provide the operator with a display whose cues match those of the human visual system. This implies drastically reducing the transmission of data which the operator cannot perceive.

The basic principles of the human visual system to be exploited include:

- High central resolution, low peripheral resolution in a wide field of view
- High spatial resolution of the position of sharp contrast edges
- Low spatial resolution of color
- Perception of depth from optic flow
- Constant fast eye motions to points of interest

The technical approach proposes the following three stages of compression:

- Resampling video data over the field of view at resolutions which match the perceived peripheral drop in human visual resolution
- Separating the data into color and edge channels
- Applying simple data compression algorithms on channel data

The result is 1600-to-1 compression ratio over raw color video transmission, and a superior field of view.

Perceptual principles and implementation algorithms are described below.

5.1 Human Visual Factors

The perceptual distribution of resolution over the human visual field, and the separation of perceptual features into channels is described in this section.

Most notably, contrast edges, which are perceived at high spatial resolution, require minimal parametric data per sample element. On the other hand, color elements, which require more parametric data per sample element, are perceived at much lower spatial resolution. Communication efficiency is gained by separating these two perceptual channels and transmitting data at rates corresponding to perceivable elements.

5.1.1 Peripheral Decrease in Resolution

The visual perception of any local elementary feature, such as a contrast edge or color, is filtered by retinal neural networks through a local receptive field which summarizes the outputs of the collection of contributing photosensors within that field into a single output. The sizes of these receptive fields differ depending on feature type, but for any fixed feature type, receptive fields increase approximately linearly in size from the center of the visual field. The process starts at the outer boundary of fovea (1 degree in diameter) and proceeds outward. At 90 degrees out, roughly the limit of the field of view, the diameters of receptive fields are more than 100 larger than in the fovea. Within the fovea, receptive fields are uniform in size, providing the highest resolution perception in the visual field.

The peripheral decrease in resolution outside the fovea can be simulated in digital imagery by applying low-pass local filters whose spans are proportional to distance from the center of the visual field. **Figure 5-1** illustrates the appearance of text without such filtering. **Figure 5-2** illustrates the appearance with such filtering corresponding to human parameters. Note that only a word or two around the center of view is readable. In **figure 5-3** the center of view shifts to another region. One can verify this rather counter-intuitive result by staring at a particular word in a line of text (such as this report) and attempting to read adjacent words without moving the eyes. By concentrating, one can discern a few words, but more than two lines away words are virtually unreadable for normal viewing distance.

fluently, including
Since Hungarian
me linguistic roots
year languages, I
e a head start in
as, it didn't work
00 Chinese charac-
le to converse with
e teacher after a
ot become fluent
my coworkers in
isappointment.

There are the Nippon
chi, and Fujitsu, as
probably Toshiba, I

In addition, there
the Sigma project,
tion of major cus-
houses. The effort
of dozen participat-
100. It is being star-
as the basis for dev-
tying the participa-
network. The part

Figure 5-1. Close-up of Text Near the Fovea

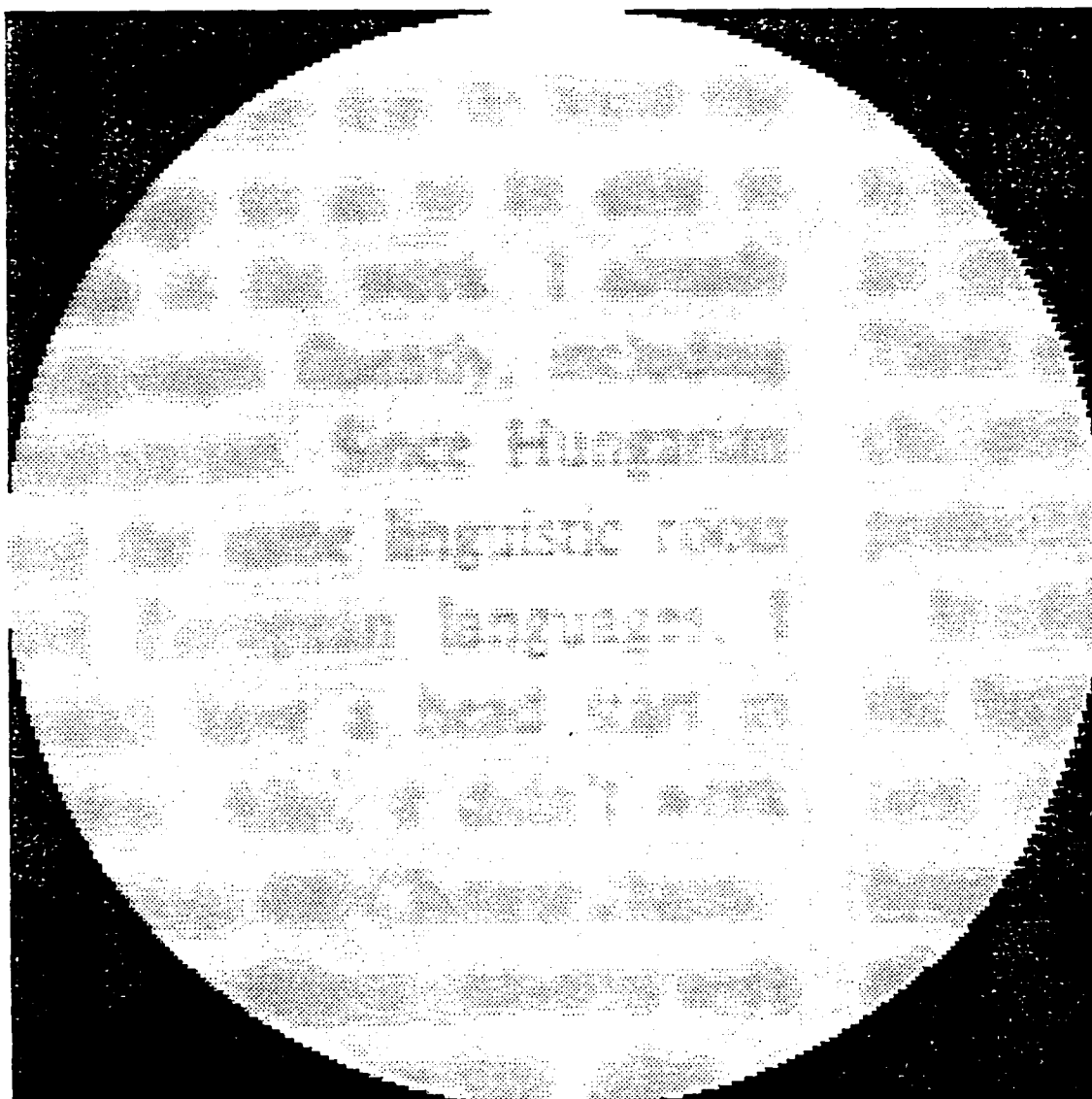


Figure 5-2. Text Showing Decreasing Perceived Resolution Peripherally

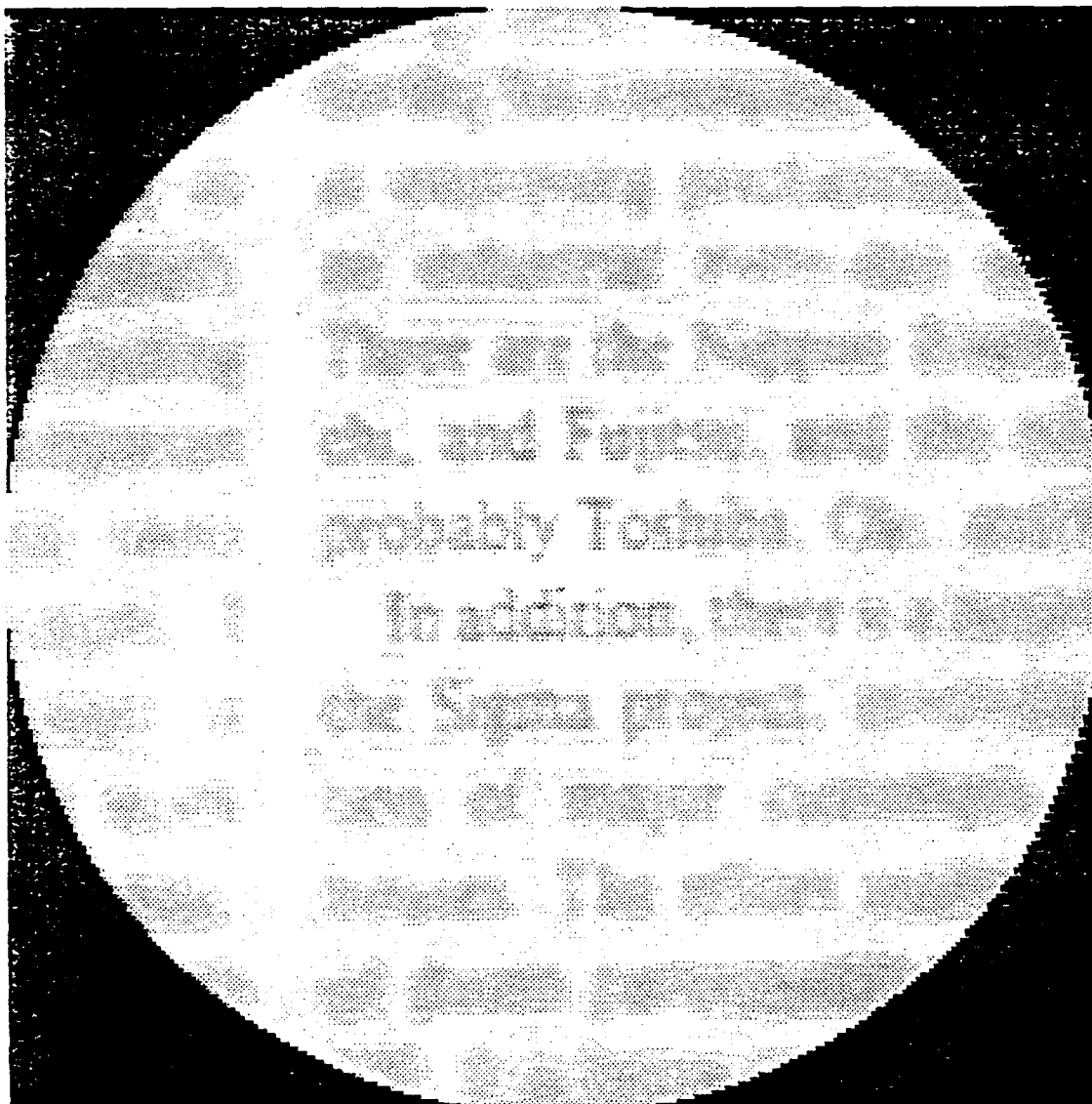


Figure 5-3. Text With Moved Center of Fixation

Another experiment, relevant to remote driving of vehicles, is instructive in demonstrating the dropoff in visual perception towards the periphery of the human field of view. Drive down a divided highway at normal speed (say 50 mph) and fix your eyes on the licence plate of the car ahead of you. At sixty feet, the fovea is viewing a span roughly one foot across, which is about the size of a license plate. You can drive quite effectively with eyes fixed on the car ahead, clearly reading the numbers on the plate while accommodating changes in speed, cars passing in other lanes, and the flow of the road and scenery around gentle curves without moving your eyes from the license plate ahead. However, it is impossible to read the plates of adjacent cars, or road signs while your eyes are fixed on the license plate ahead.

Another observation relevant to driving is the annoyance experienced when following a panel truck which obstructs the center of view. A six foot square back panel at sixty feet occupies roughly 1% of the total field of view (10% in each dimension), but because of the distribution of receptive fields the image of that back panel occupies slightly more than half of the data representation area in the visual cortex. Furthermore, it obstructs the center of motion towards which the eye instinctively is drawn, as the source of impending events.

The peripheral decrease in resolution is discussed below in terms of the two primary perceptive channels, contrast edges and color.

5.1.2 Brightness Contrast Edges

Contrast edges, local boundaries between dark and light regions, are the primary cue in human visual perception. Our binocular and monocular depth perception, and recognition of patterns is critically dependent on our ability to accurately detect and locate contrast edges. The use of cartoons, sketches, blueprints, aircraft silhouette charts, signage, and typefonts for instantaneous visual communication confirm a pervasive human visual talent to perceive complex patterns from limited edge-based data. The fineness at which we can perceive edges is a measure of visual resolution known as acuity.

A commonly used measure of human visual acuity is the ability to detect fine texture patterns consisting of alternating dark-light bands. Measurements of acuity can be made by presenting ripple patterns consisting of sinusoidal brightness waves. Acuity is defined in terms of the finest perceivable texture.

Figure 5-4 illustrates the spatial frequencies detection envelope for humans in the fovea. Higher modulation (contrast) yields higher acuity at spatial frequencies above five cycles per degree (i.e. 12 arc minutes of the field of view). The temporal axis illustrates that flickering the image increases the ability to discern low spatial frequencies (broad textures). The maximum spatial frequency discernable under extreme contrast lighting conditions is 20 cycles per degree, or 3 arc minutes per cycle. Video contrast does not ordinarily provide such conditions, particularly when navigating in natural environments. A more reasonable acuity limit for viewing video is 6 arc minutes per cycle. In these terms, one cycle consists of a contrasting dark and light region resolvable within the field of view. This requires a diameter of two pixels for display. Hence minimal pixel resolution required to match such a rate is 3 arc minutes, which is approximately the angle subtended by 1 inch at 30 yards.

Experiments and neurophysiological studies show that the human visual field is divided into small processing regions, called receptive fields, each of which provides summary information to the brain. The size of a receptive field is roughly double the magnitude of visual acuity. That is, one contrast cycle from light to dark or vice versa (i.e. two brightness values) can be perceived by a receptive field.

The human visual field actually reaches to over 90° on the outboard side, and about 45° nasally, but a viewing field of 90° covers enough to provide excellent peripheral vision. At the center of the field of view there is a uniform high acuity region about 1 degree in diameter, called the fovea. Within this disk maximum acuity is about 3 arc minutes under optimal lighting conditions. The size of receptive fields is uniform within the fovea, and grows linearly with retinal distance from the center outside the fovea. **Figure 5-5** illustrates this trend in graph form. **Figure 5-6** depicts the distribution of receptive fields.

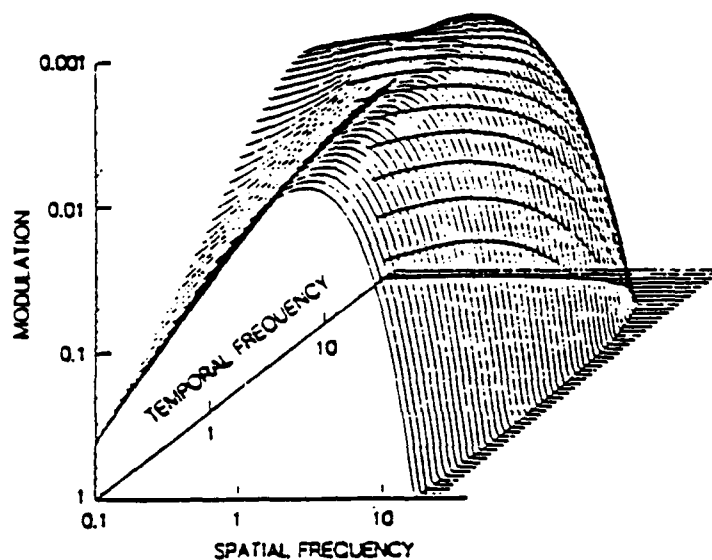


Figure 5-4. Spatio-Temporal Resolution Envelope for Humans

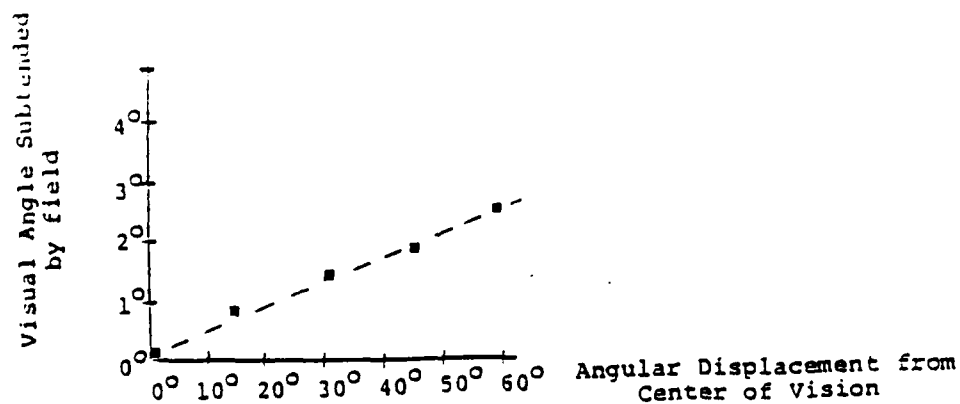


Figure 5-5. Human Receptive Field Diameters

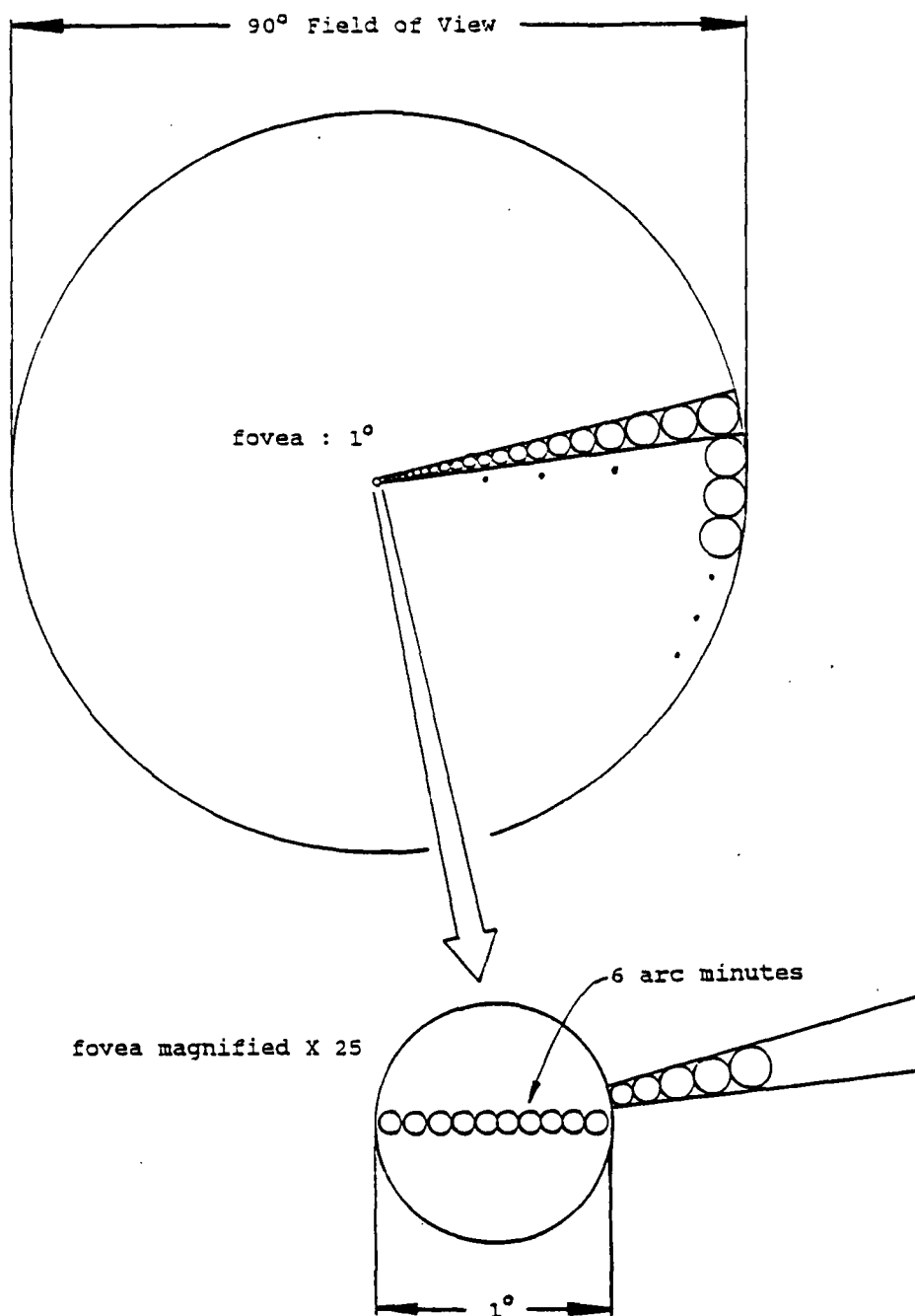


Figure 5-6. Human Field of View and Receptive Field Distribution

5.1.3 Color Channel

Color vision provides humans with a competitive edge over colorblind creatures in nature in unmasking camouflage, providing instant wide view context, and complementing depth perception. These capabilities are very useful for remote driving, visual targeting, and threat detection.

Color vision is based on three types of photoreceptive cells (cones) in the retina which are spectrally sensitive in the red, green, and blue parts of the spectrum. However, before these are transmitted to the brain, they are combined by "opponent process" neural networks whose outputs are the differences between outputs of the neighboring cones of differing colors. The result is a trichromatic perception of color which is independent of brightness. That is, any color can be matched by a linear combination of three primary colors, and scaling the coefficients in this linear combination increases the brightness but leaves the perception of color unchanged. These and other properties of color processing in neural networks in eye and brain lead to a perceptual space for color which organizes image elements into two color dimensions and one brightness dimension, as illustrated in **figure 5-7**. The first color dimension, hue, corresponds to the intuitive notion of position in the color spectrum or on a color wheel. The second, saturation or chroma, corresponds to the clarity, or lack of "muddiness" of the color.

There are numerous coordinate systems for color, based on needs to define paint pigments (Munsell system) establish international standards (CIE), or compartmentalize perceivable color differences in psychophysics. In computer graphics and image processing, the RGB system represents color as the independent intensities of red, green and blue primaries. The advantage is that color components can be treated computationally exactly as brightness in filtering and interpolation algorithms. The disadvantage for image transmission is a tripling of bandwidth.

The local "opponent process" networks in the retina operate over receptive fields which are up to five times larger than edge detecting receptive fields. That is, color acuity is much lower than edge acuity. This properties will be exploited in the color coding scheme for image compression described in section 5.3.1.

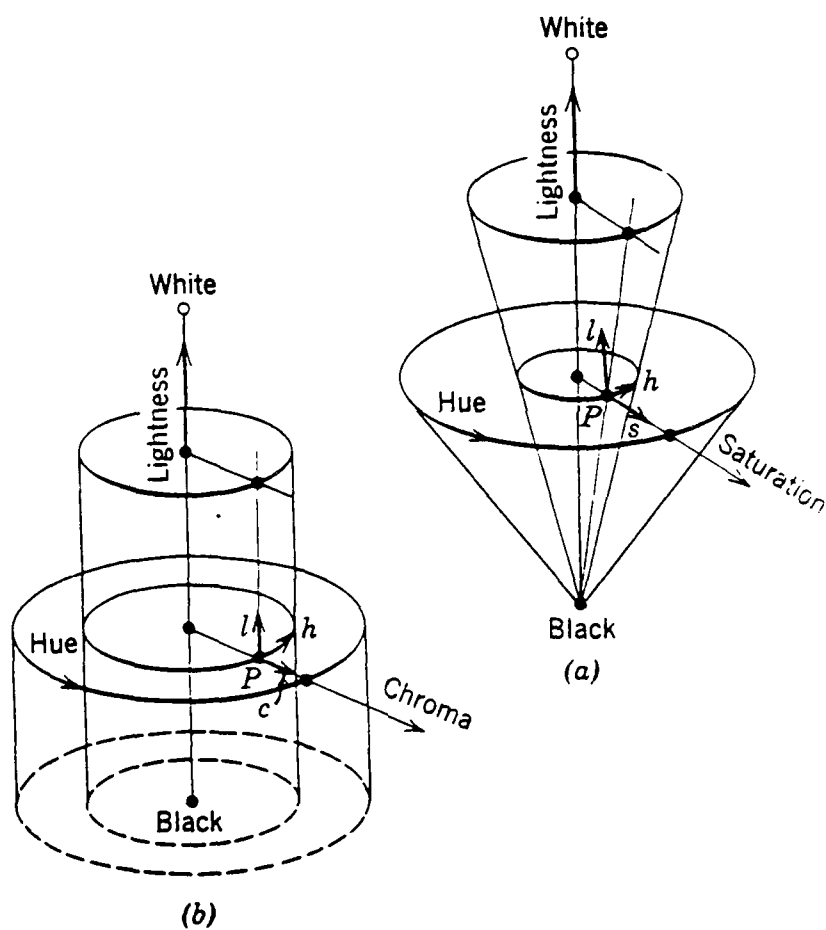


Figure 5-7. Color Perception Parameter Spaces

5.1.4 Implications of Visual Acuity for Display Resolution

Both RS-170 (monochrome) and NTSC (color) analog video signal standards specify 525 lines per video frame. Digital frame grabbing, image processing, and robotic vision often reduce this to 512, of which 480 are displayed between blanking periods, because 512 is a power of two and therefore more amenable to digital addressing. A pixel is designated as 1/512(th) of a video line. That is, the analog line signal is broken into time slices which are digitized, typically to 8 bits each for brightness in monochrome, or 8 bits per color for color. Each pixel defines a geometric tile on a display screen which corresponds to a viewray in the field of view. The size of a pixel is the ultimate resolution unit for cameras or displays.

A typical CCD camera, color or black and white, contains a chip of dimensions 8.8 by 6.6mm (.34 by .25 inches), referred to as 2/3 inch format. The widest field of view typically available without fisheye distortion is achieved with a lens of focal length 8 mm. This corresponds to field of view 57 degrees wide and 42 degrees high. Pixel width for a 512 pixel line then corresponds to .11 degrees or 6.6 arc minutes. Thus 8 millimeter focal length closely matches resolution of the human eye at the fovea. Longer focal lengths yield proportionally narrower fields of view and correspondingly finer angular resolution per pixel, implying overresolution.

Display screens are nominally flat and designed to be viewed from a direction perpendicular to their surface. Thus perceived pixel size is uniform across the field of view and varies with viewing distance. Normal viewing distances for television are proportional to screen size. For any particular size, viewing too close yields a grainy perception of scan lines and blobs of light at pixels. In viewing from too far, details are missed.

For teleoperation, consider a display which is at the proper distance from the operator to faithfully subtend the same field of view as the camera. For example, a 5 inch diagonal screen with 3:4 aspect ratio is 4 inches wide. To subtend 57 degrees as in the preceding camera example, the display must be about 3.6 inches in front of the viewers face. Similarly, a 21 inch display must be 15 inches in front of the viewer. For broadcast television, most focal lengths are much longer, reducing the field of view and increasing the corresponding distance from viewer to screen.

Figure 5-8 illustrates the relation between screen size and viewing distance to match camera and operator field of view for a 16mm lens, which subtends a field of view width of 30 degrees. This puts a 5 inch display at just under 8 inches from the viewer, a 13 inch display at 20 inches, and a 21 inch display at 31 inches. One pixel in a 512 pixel line subtends .05 degrees or 3.5 minutes of arc, adequate for maximum acuity display.

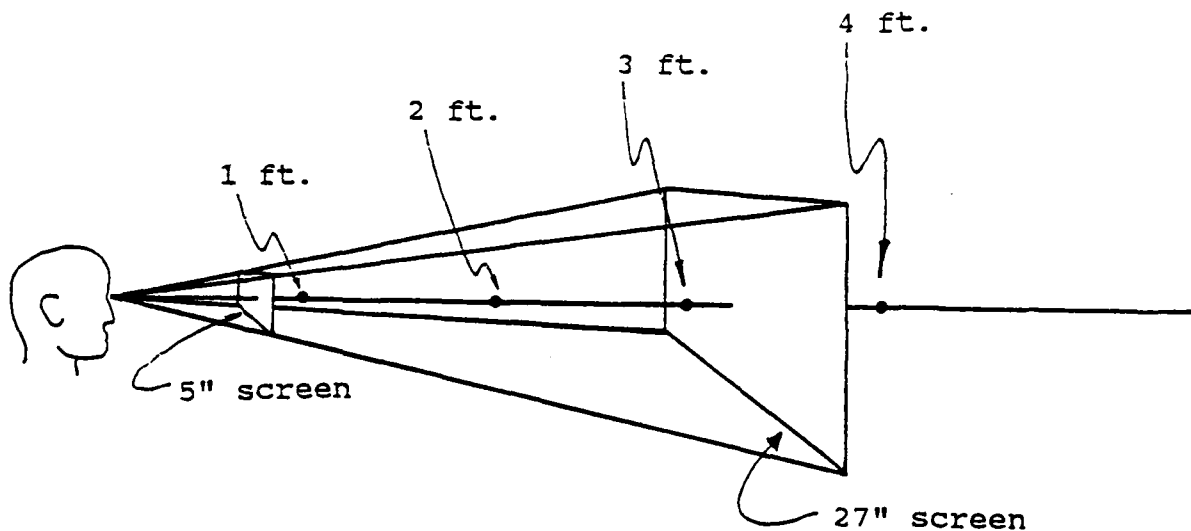


Figure 5-8. Nominal Viewing Distance for Video Display Screen

Figure 5-9 superimposes most of the human visual field on a video display at typical viewing distance. The human visual field actually reaches to over 90° on the outboard side, and about 45° nasally, but a total viewfield of 90° covers enough to provide excellent peripheral vision. The screen occupies 30 degrees of width in the field of view, corresponding to a 16 mm camera lens. Note that the fovea occupies approximately 3% of screen width or about .1% of screen area. Note also that the human field of view far exceeds that of the screen or camera. This is bad for remote vehicle operation. Many depth perception cues arise from motion in the peripheral field of view, which is not captured by most cameras and lenses. Width of field of view and viewfield resolution are necessarily conflicting requirements. A telephoto lens (long focal length) magnifies a small region exactly proportional to the restriction in field of view. A display resolution gradient can provide the best of both worlds, namely, a wide field of view with high detail in the center.

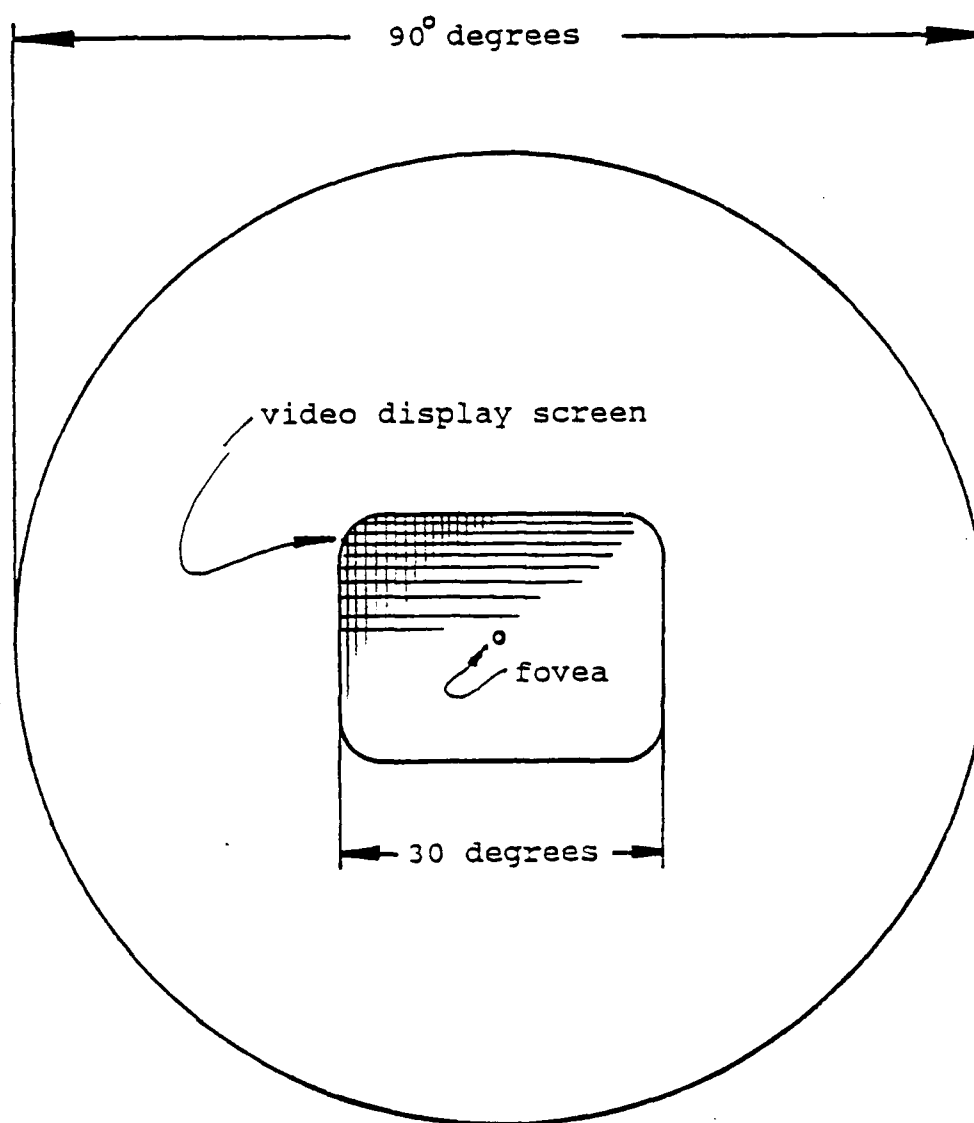


Figure 5-9. Video Display Screen within Human Field of View

5.2 Logarithmic Mapping and Reconstruction

The radial drop in acuity from the center to the periphery of the human visual field can be incorporated into image acquisition algorithms which efficiently resample video data, rearranging the desired rotational sampling pattern into a memory array. This mapping is known mathematically as the (conformal) logarithmic mapping. Details may be found in [Weiman 88].

Methods of implementing the logarithmic mapping, and reconstructing imagery which can be usefully perceived by the operator, are described below.

5.2.1 Logarithmic Mapping

The radial decrease in human visual acuity illustrated in **figure 5-5** can be represented as a rotationally symmetric sampling pattern whose cells increase in size linearly from the origin. Each cell corresponds to a visual receptive field. In mammals, the multitude of photosensor cells (typically a few hundred) within each receptive field contribute to a summarized response transmitted as a unit to the brain. The largest component of data compression in our technical approach is based on capturing and transmitting data which matches this distribution of human perceptive fields.

The sampling pattern in **figure 5-6** can be implemented computationally by transforming x-y coordinates of video to $\log(r)$ -theta (log polar) coordinates, that is,

$$\ln(r) = \ln(\sqrt{x^2 + y^2})$$

$$\theta = \arctan(y/x)$$

Quantizing r and θ , and summing all x-y data items within each quantum range reduces all data within each such receptive field single element. The reduced data may be stored into an array whose rows and columns correspond to the rays and rings of the rotationally symmetric sampling pattern of cells, as illustrated in **figure 5-10**. This mapping is well known mathematically as the conformal logarithmic mapping, denoted by $\text{LOG}(z)$.

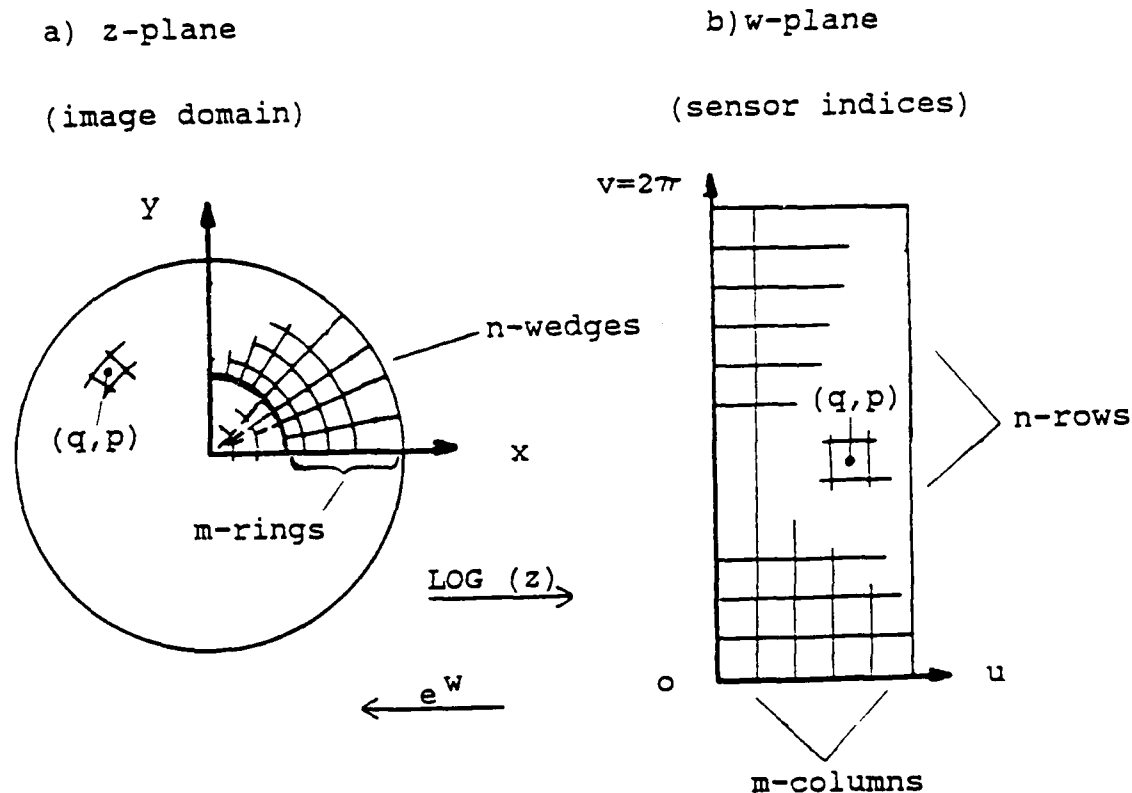


Figure 5-10. Conformal Mapping for Resampling Video Imagery

Software implementation of the logarithmic mapping averages the data within each sampling cell, and dumps the result in the (u,v) array. The resolution of the sampling array corresponds to its number of rays, which in turn corresponds to the number of rows in the u-v array. The field of view of the sampling array corresponds to its number of rings, which in turn corresponds to the number of rings in the u-v array. If "n" is the number of rays in the sample space, the number of rings is

$$m = [\ln(r_{\max}/r_{\min})] * n / (2\pi)$$

where r_{\max} and r_{\min} are the inner and outer radii of the sampling pattern. Details are given in [Weiman 88a].

One can make the central disk, which is not mapped, as small as desired. One pixel diameter is the limit for processing, but a perceptually reasonable limit would be a diameter which somewhat exceeds the human fovea.

That is, since human resolution is uniform within the fovea, one gains nothing by displaying data at higher resolutions. In practice, a diameter of two times the fovea permits the eye to drift somewhat on the display without losing the perception of uniform resolution.

Consider the reduction in data transmission using the log polar mapping in the example of figure 5-9. The central 1 degree of the 512 pixel display lane is 17 pixels wide. The area of the disk is thus 228 pixels. Using the human acuity measure of 3 arc minutes implies about 64 pixels per ring in a log polar resampled image (as in figure 5-10). The number of rings is

$$q = (n/(2\pi)) \cdot \ln(r_{\max}/r_{\min}) = 34$$

where $n = 64$ resampling wedges, $r_{\min} = 9$ pixels and $r_{\max} = 256$ pixels. Thus the total number of pixels in the mapped display which matches human acuity is:

$$npix = 228 + 34 \cdot 64 = 2404.$$

The ratio of original pixels to mapped pixels is thus

$$512 \cdot 512 / npix = 109:1.$$

That is, two orders of magnitude reduction are achieved simply by mapping. This number is roughly four times the compression rate (25:1) expressed in the example in section 4.1.4, as a result of the doubled resolution there ($n = 128$) required to detect edges in the proposed field system. That is, doubling resolution quadruples resampling cell count.

Hardware implementation of the logarithmic mapping follows scan-line order, mapping each frame of data on-the-fly. Lookup tables route data to appropriate destinations. TRC has fabricated an IBM PC plug-in board which logarithmically maps 256-by-256 pixel black and white video images at 30 frames per second. Incoming pixel addresses route pixel data to bins which correspond to the resampling cells via lookup table. TRC has also developed software for mapping color and black and white images at a few seconds per frame on an IBM resident frame buffer commercially available from Imaging Technologies Inc.

Real-time logarithmic mapping can also be accomplished on the NASA - Texas Instruments Remapper. The large size and power consumption of this equipment are prohibitive for mobile applications in, for example, the RCC. The total enclosure occupies more than five cubic feet, weighs about two hundred pounds, and consumes about one hundred and fifty watts of power.

5.2.2 Inverse Logarithmic Mapping

Photo 6 illustrates a typical raw scene together with its logarithmically mapped image in the upper left hand corner. The size of the rectangular region relative to the size of the original image corresponds to the compression ratio resulting from the mapping. However, the distortion of the geometry of mapped data renders it useless for human interpretation in that form. It must be reconstructed to replicate the geometry of the original display. But the geometric compression of the logarithmic mapping replaced sampling regions with individual points. The missing display points must be filled in by interpolation.

TRC has developed a highly efficient algorithm for image reconstruction based on Gaussian Interpolation. This technique was generalized from a method used for 2-by-2 interpolation of image points in image pyramid expansion, reported in [Burt 83]. The TRC generalization permits indefinitely large neighborhood sizes without increasing the computation required. Software implementation of the Gaussian Interpolation algorithm has been used to generate photo 4, using less than a minute of computing time on an IBM PC.

Figure 5-11 illustrates the geometry of reconstructing display data from logarithmically mapped data. Source data in the log domain (part b) consists of samples L in a (q,p) array, represented by small triangles. These are to be mapped into a display array as shown in part "a". Display scanline coordinates (x,y) contain more data points (small squares) than the source. Thus, source data must be interpolated to generate display data.

The simplest possible reconstruction method, zeroth order interpolation, consists of replicating $L(q,p)$ for each (x,y) within the image of the domain of the source cell. The result is large tiles of uniform value bordered by step discontinuities, as shown in photo 7. To provide seamless reconstruction, it is necessary to smoothly interpolate data according to its position among the samples of the data point array.

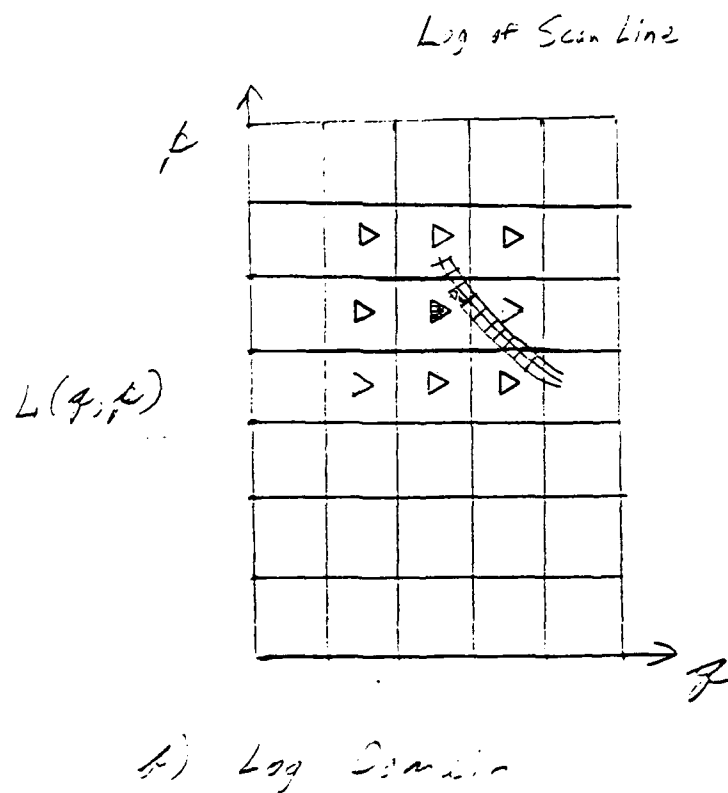
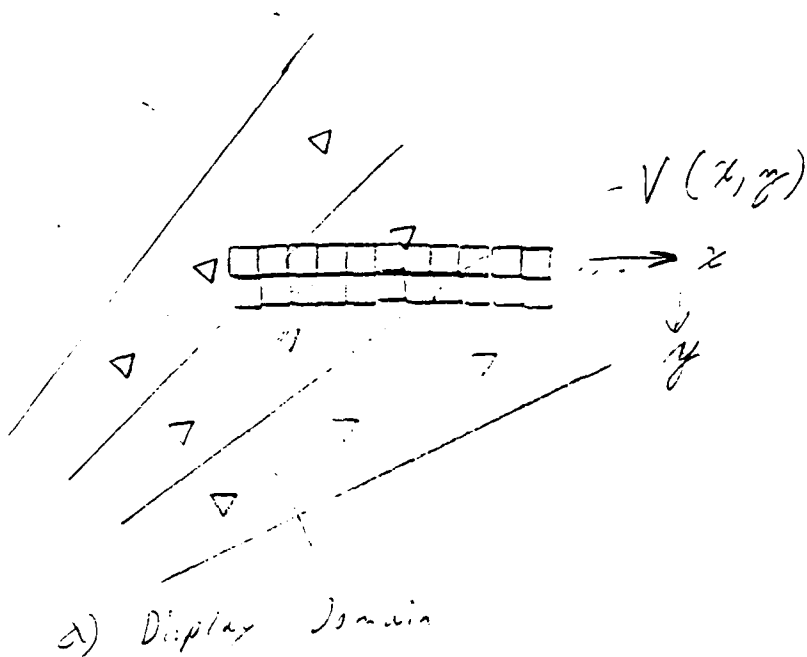


Figure 5-11. Geometry of Image Reconstruction Interpolation

Gaussian interpolation involves filtering (convolution) sparse source data with a 2-D Gaussian whose bandpass closely matches that of the source data. It can perform as well as bicubic interpolation with far less computation.

Figure 5-12 is a blow-up of figure 5-11b, illustrating the generation of a single x-y pixel from the 3x3 nearest neighbors in the logarithmic domain. The log image of the x-y pixel is indicated by the small cross-hatched square. The small triangles represent data points in the Log array, centered in large square cells which represent their domains. The concentric circles centered on the small square represent the one-, two- and three-sigma radii of a 2-D Gaussian function centered on the small square.

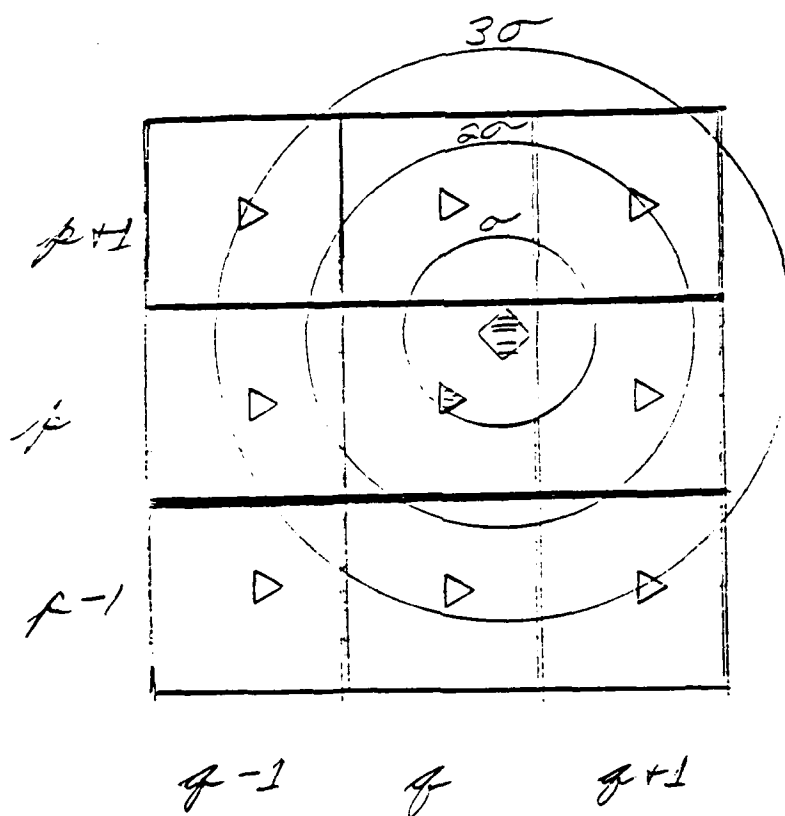


Figure 5-12. Geometry of Gaussian Interpolation

Gaussian interpolation consists of summing the Gaussian weighted values of each input pixel in the 3x3 neighborhood of cell (q,p). Weights consist of the volume under the Gaussian surface, partitioned by the (q,p) cell boundaries. Thus, the equation for the value at output pixel (x,y) is

$$V(x,y) = \sum_{i=-1,0,1} \left(\sum_{j=-1,0,1} W_{ij} * L(q+i,p+j) \right)$$

where W_{ij} 's are the volumes under the Gaussian cookie-cuttered by the (q,p) cell boundaries.

Implementation of Gaussian Interpolation Since the 2-D Gaussian is separable, its values are the product of q and p components in figure 5-12. This reduces lookup table size by orders of magnitude. That is,

$$\begin{aligned} g2d(u,v) &= (pi^{*-1}) * exp(-(u^{*2} + v^{*2})/2) \\ &= (pi^{*-.5}) * exp(-u^{*2}) * (pi^{*-.5}) * exp(-v^{*2}) \\ &= g(u) * g(v) . \end{aligned}$$

Figure 5-13 illustrates the separation of the Gaussian along two axes. Thus the weight for each contributing pixel is the product of the weights along two components, i.e.,

$$W_{ij} = w_i * w_j'$$

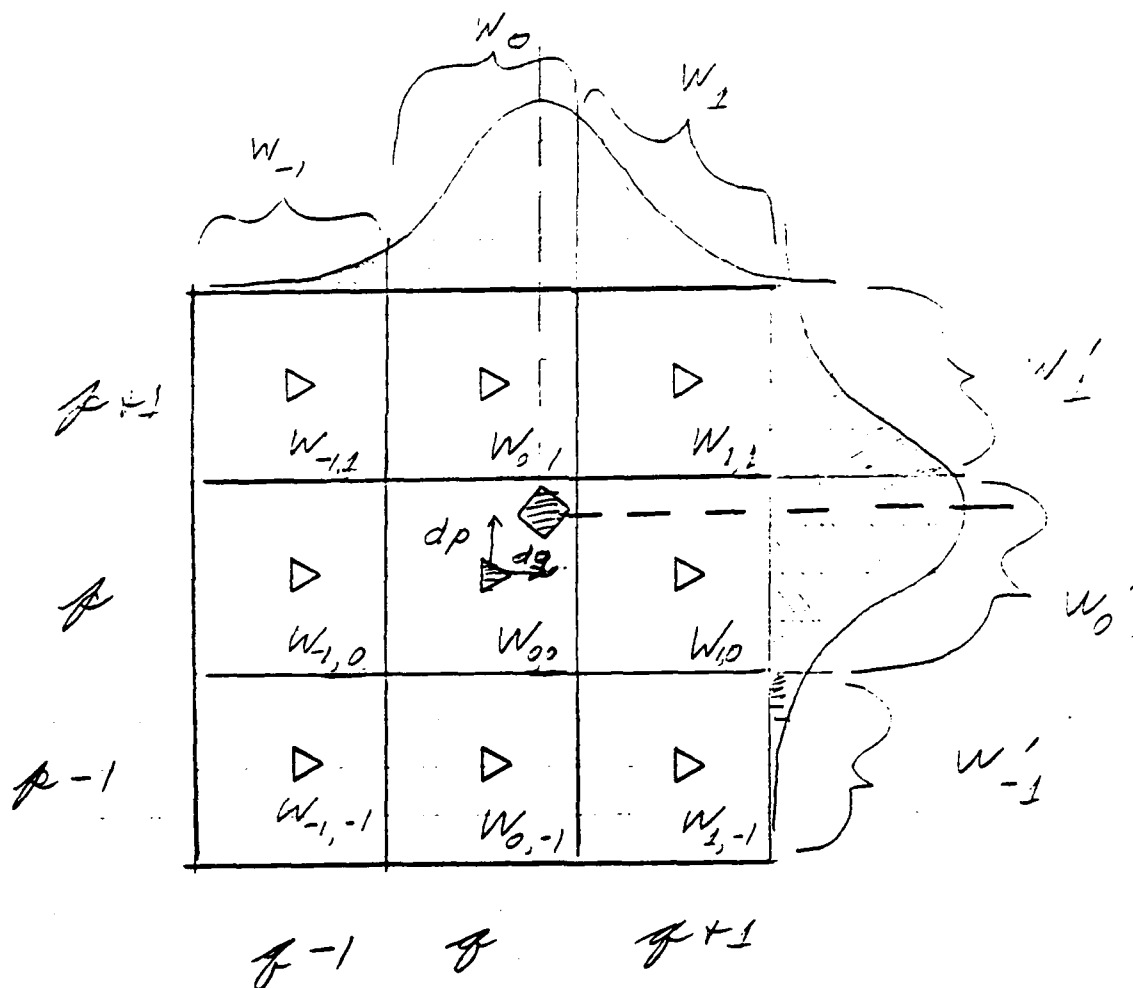


Figure 5-13. Separation of 2-D Gaussian into two 1-D Gaussians

The offset of the target cell (hatched square in figure 5-13) within the 3x3 neighborhood of source cells can be represented by subpixel addresses (dq,dp). These fictitious locations are used to offset the lookup into the Gaussian table as follows.

Figure 5-14 illustrates the integral of the one dimensional Gaussian,

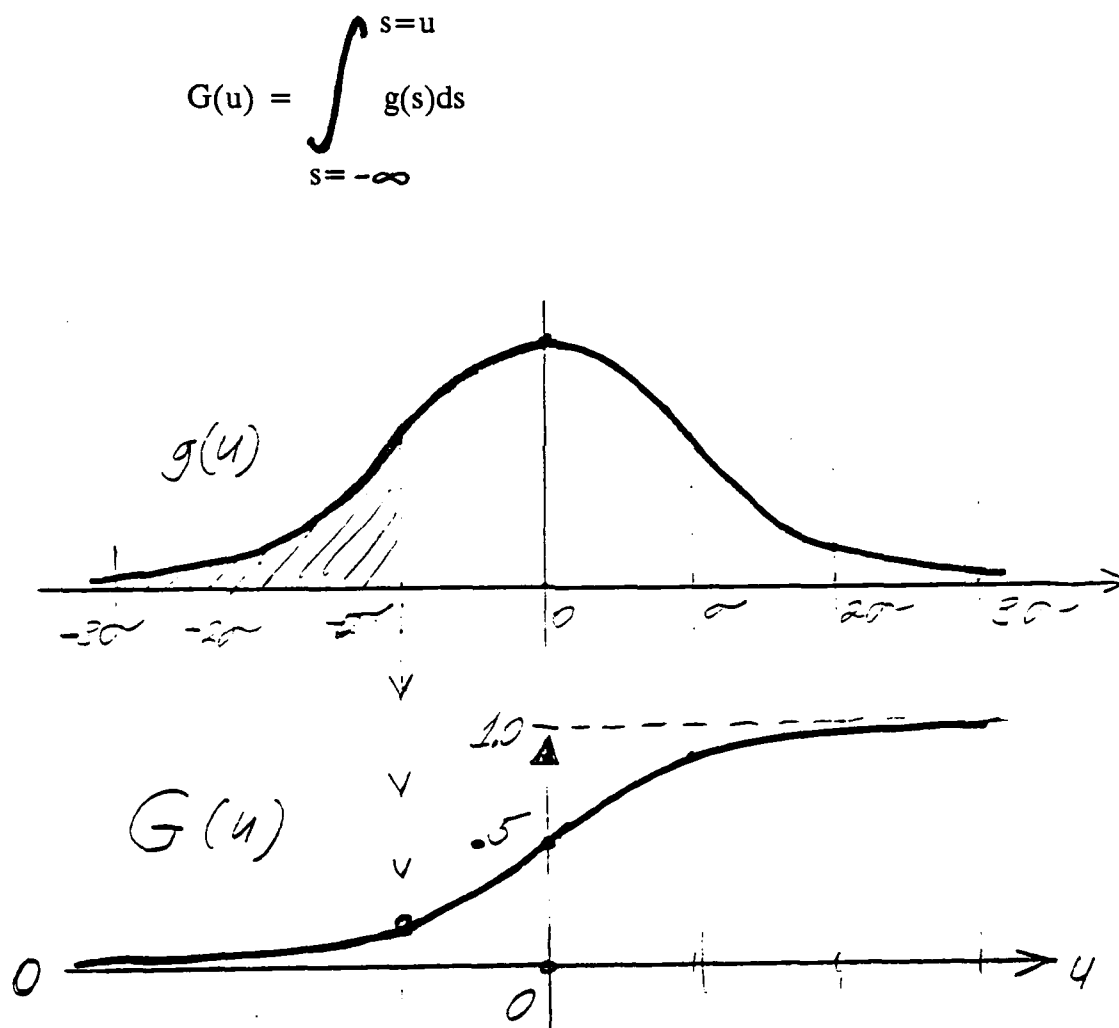


Figure 5-14. Integral of 1-D Gaussian

The weights, which are areas under the curve over the intervals in question, are computed as follows.

$$w_{-1} = G(-1-2dq) = GLT_{-1}(dq)$$

$$w_0 = G(1-2dq) - G(-1-2dq) = GLT_0(dq)$$

$$w_{+1} = 1 - G(1-2dq) = GLT_{+1}(dq)$$

The notation $GLT_i(dq)$ reflects the property that the expressions may be stored as fixed one parameter lookup tables for efficient computation.

The scaling choice of 1/2 pixel per sigma appears optimum to prevent aliasing while maximizing information transfer to the resampled grid. That is, the one sigma cutoff in the frequency domain corresponds to a wavelength which is $\pi/2$, or roughly 1.5 times the Nyquist rate of the source grid. Energies at frequencies higher than that ought to be below 2 percent. Slightly tighter sigmas could be used at the risk of introducing aliased high frequency tile boundaries.

Realtime hardware implementation is straightforward. Scanline ordered lookup tables point to the logarithmic mapped data. Subpixel bits correspond to dp and dq which assign gaussian weights.

Real-time reconstruction can also be accomplished on the NASA - Texas Instruments Remapper. This general purpose mapper performs 4-by-4 interpolation on pixels at video rates.

5.3 Edge and Color Coding and Reconstruction

The preceding section described methods for logarithmically resampling an image, and reconstructing the data. This section describes methods for coding and reconstruction of the perceptual content of mapped imagery.

5.3.1 Color Coding and Reconstruction

In [Faugeras 76], maximum color acuity is given as 4 cycles per degree for red-green contrast and 2 cycles per degree for yellow-blue contrast. These correspond to 15 arc minutes and 30 arc minutes of color acuity, which are 2.5-to-1 and 5-to-1 lower than contrast edge acuity. Picking 4-to-1 as a conservative average, color data may be mapped by choosing resampling cells four times as large as edge resampling cells. This results in a 16-fold (4-by-4) reduction in resampling cells, a significant improvement for bandwidth compression.

Photo 8 illustrates the reconstruction of a mapped color image at twice the resolution of **photo 2**. The processing for all color photos was carried out on a low resolution (256 line) color frame buffer (ATT Image Capture Board) which separates red, green and blue components in fields within a 16 bit word. Logarithmic mapping and reconstruction were implemented by Gaussian interpolation on each color component separately.

Color data compression rates could be doubled by encoding color according to perceptual spaces as illustrated in **figure 5-7** rather than RGB. By categorizing source pixels into 8 basic spectral hues, 4 levels of saturation, and 8 levels of brightness, color resampling cells could be characterized by 8 bits each instead of 16, prior to transmission. These color sample cells could be optimally chosen to match chromaticity compartments scaled to human color discrimination resolution, illustrated in **figure 5-15**. This is a CIE color system diagram with the RGB gamut of color TV shown in solid lines. The fin-shaped path around the central region is the locus of pure spectral colors, whose wavelengths are annotated (in nanometers) along the boundary. That is, red is about 610 nm, green is 540 nm, and blue is 470 nanometers. The ellipses represent perceptual discriminability; more subtle shades of blue can be discriminated than green. The reconstruction process could use Gaussian interpolation with 24 bits of display color, completely eliminating artificial boundaries caused by the low number of color compartments used in the resampling process, at the expense of somewhat garish but highly discriminable color.

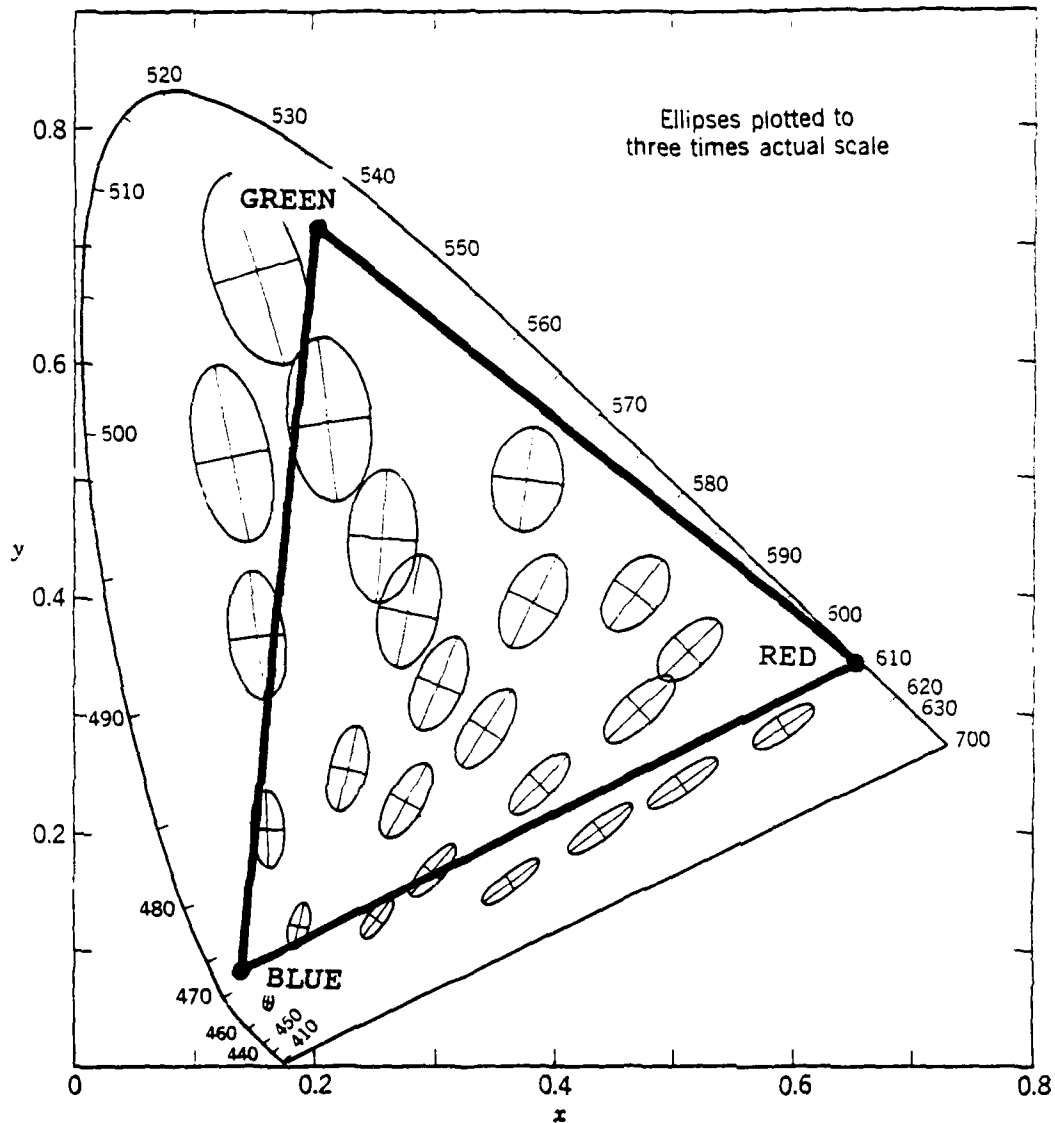


Figure 5-15. Color Discrimination Superimposed on CIE Diagram

5.3.2 Edge Detection and Reconstruction

Visual contrast edges consist of local regions with large variations in brightness. They can be computed using functions which take finite differences of neighboring pixel values. Large magnitudes signify discontinuities or edges. Such difference operations are customarily applied to images by convolving a template of weights with every neighborhood in the image.

Figure 5-16 illustrates the neighborhood weighting operations defining some common edge detectors. Most of these operate on small neighborhoods of pixels, usually 3x3. This permits them to be computed at video rates with simple image processing hardware, but may limit their noise rejection properties.

Figure 5-16 is to be interpreted as follows. Each square is a template of coefficients which multiply pixel contents of the region of the image upon which the template is superimposed. the output is an edge image array derived from the original image array. For example, the output of the Roberts operator for pixel (x,y) is

$$R(x,y) = \text{abs}(H_1) + \text{abs}(H_2) = \\ \text{abs}(f(x,y+1) - f(x+1,y)) + \text{abs}(f(x+1,y+1) - f(x,y))$$

where "abs" is the absolute value and $f(x,y)$ is the contents of the (x,y)th pixel. The array $R(x,y)$ will contain large values where the underlying pixels of $f(x,y)$ exhibit local contrast.

The Sobel operator has components H_1 and H_2 which are brightness gradients in the x and y directions. The ratio of these is the tangent of the angle of orientation of the underlying edge, and the square root of the sum of the squares is the magnitude of the edge gradient.

Roberts

$$H_1 = \begin{array}{|c|c|} \hline 0 & -1 \\ \hline 1 & 0 \\ \hline \end{array} \quad H_2 = \begin{array}{|c|c|} \hline -1 & 0 \\ \hline 0 & 1 \\ \hline \end{array}$$

Prewitt

$$H_1 = \begin{array}{|c|c|c|} \hline 1 & 0 & -1 \\ \hline 1 & 0 & -1 \\ \hline 1 & 0 & -1 \\ \hline \end{array} \quad H_2 = \begin{array}{|c|c|c|} \hline -1 & -1 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline \end{array}$$

Sobel

$$H_1 = \begin{array}{|c|c|c|} \hline 1 & 0 & -1 \\ \hline 2 & 0 & -2 \\ \hline 1 & 0 & -1 \\ \hline \end{array} \quad H_2 = \begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

Figure 5-16. Edge Detection Filter Templates

Small span edge detectors such as those of **figure 5-16** often respond spuriously to random noise over small neighborhoods, and fail to detect useful trends in neighborhoods larger than three pixels in diameter. These problems are overcome by large span edge detectors such as Grossberg's, illustrated in **figure 5-17**. Rather than being square as those in **figure 5-16**, the template is oval, with positive weights above the center, negative below, and tapered values towards the boundaries. The problem with Grossberg's and related directional edge detectors is that a complete convolution pass over the image data yields edges in only one direction, whereas omnidirectional operators such as Roberts or Sobel detect edges of all orientations in one pass. The template in **figure 5-17** detects changes in brightness in the vertical dimension, but not in the horizontal direction. This may be desirable from the point of view of data reduction, but may cost the user some useful data. In driving, horizontal and vertical edges are important. The direction of gravity is a reference for motion and surface. Cat's eyes have vertical pupils which gives them preferential optical resolution for horizontal lines, which might represent prey profiles protruding over the horizon of cover.

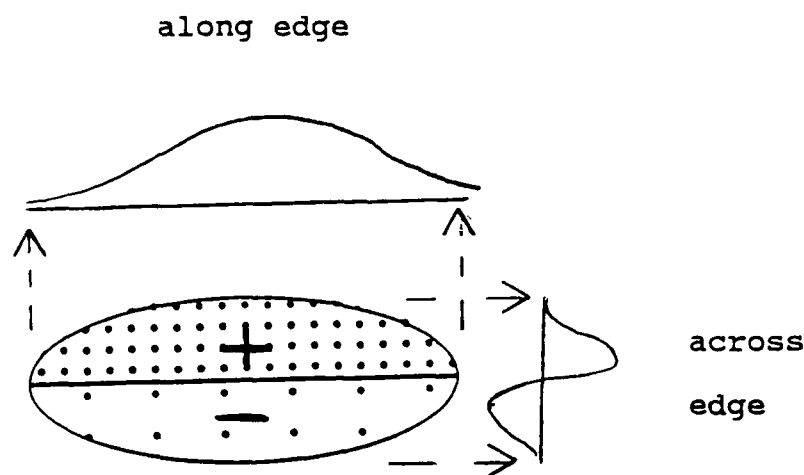


Figure 5-17 Grossberg's Directional Edge Detector

In the proposed image compression system, the second stage operates in the log domain. The resulting reduction in pixel count by at least an order of magnitude correspondingly reduces image processing rate requirements. Color coding requires several bits per pixel using a low resolution resampling pattern. Since edges can be coded as present or absent, only one bit per pixel is necessary, an advantage for image compression.

Gaussian interpolation provides high quality reconstruction of sparse data at minimal computational cost without aliasing artifacts. Its efficiency derives from display scanline ordered computations over minimal (3-by-3) neighborhoods. However, it does not reconstruct edges with appropriate visual cues. Its very smoothness, which arises from interpolation of sparse source data into dense display pixels, dulls edges which should be displayed sharply, as shown in **Photo 9**, which resulted from the following process. Black and white data in **photo 1** was logarithmically mapped and then subjected to Sobel edge detection. The result was inverse mapped using Gaussian interpolation to restore the original image geometry. Edges which are sharp in the log domain are smeared into blobs as shown in **photo 9**. The cross section of such blobs is a hump, illustrated in **figure 5-18a**.

There are a number of alternatives for sharpening dull display edges which come through gaussian interpolation. The simplest is to threshold the hump as shown in **figure 5-18b**. The problem with this approach is that peripherally displayed edges, which are wide blobs, will come through as very wide stripes, obliterating scene features and giving false cues as regions rather than boundaries.

Another approach to sharpening reconstructed edges is to find the ridgelines of the humps. Mathematical morphology can be used to thin these features until only a skeleton is left. A drawback is that many iterations are required for the extended humps in the periphery, fatally delaying realtime display.

Alternatively, local maxima could be used to define ridge lines in three pipelined steps, illustrated in **figure 5-18**. A gradient operator such as the Sobel is applied to the hump, yielding **figure 5-18c**. **Photo 10** illustrates the operation of this process on the data of **photo 9**. The desired output in **figure 5-18d** results from setting to 1 those pixels for which exceed some threshold T1 in part "a", provided the corresponding pixel in part "c" is less than some threshold T2. **Photo 3** illustrates edges resulting from applying such a process.

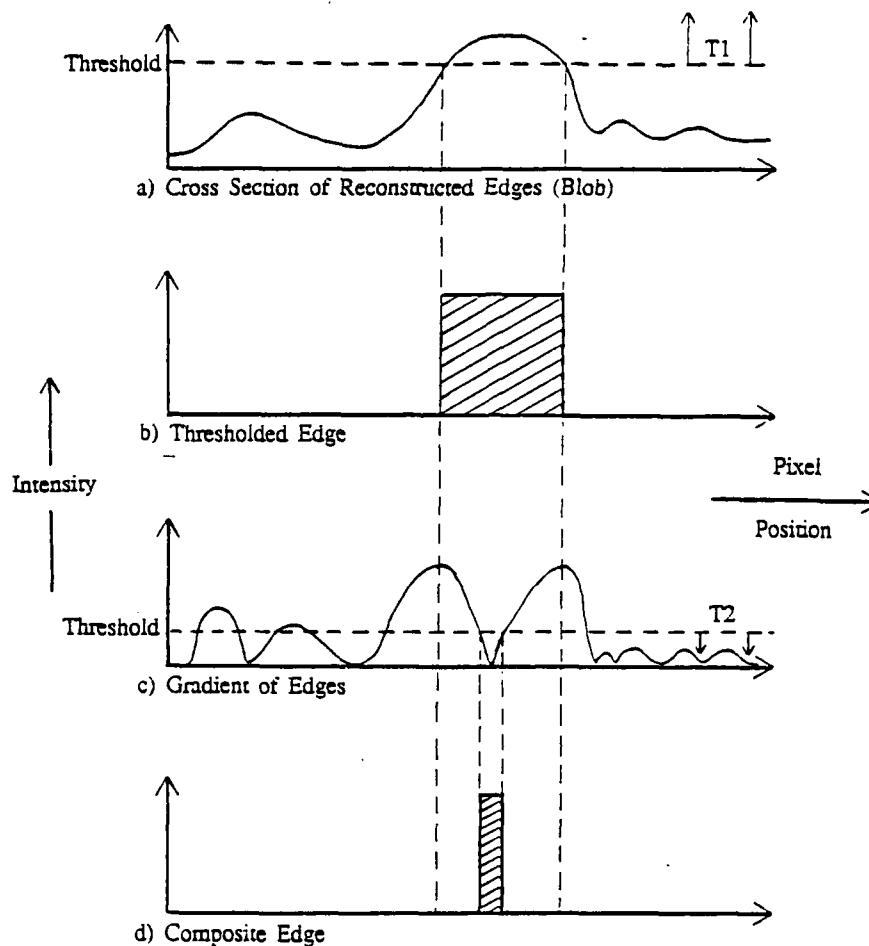


Figure 5-18. Sharpening of Edges Reconstructed from the Log Domain

The methods for reconstructing edges which were simulated in Phase I are not sufficient for remote driving. Much better results can be achieved with further experimentation. Perhaps the most promise lies in retaining linear feature data representation of edges rather than treating edges as intensity information. That is, in the log domain edges should be represented as short segments of lines or curves, characterized by control points and direction. Reconstruction would consist of mapping the control points to display space, interpolating thin curves between them rather than spreading them out with Gaussian interpolation. Since linear feature data is at least an order of magnitude sparser than pixel data, considerable improvement in data compression would also result from this approach. **Photo 11** illustrates the display concept.

5.4 Image Compression Techniques

Once color and edge information have been distilled from logarithmically mapped images, the resulting data can be subjected to standard data compression techniques, which comprises stage 3 of the image compression process. A variety of these were examined and simulated by Dr. Norman Griswold of Texas A&M University under subcontract to the Phase I SBIR. Methods analyzed include Hybrid Cosine/DPCM (differential pulse code modulation), Block Truncation Coding (BTC), quadtree coding, Karhunen-Loeve, and other techniques. Hybrid Cosine/DPCM coding is highly recommended, yielding 8-to-1 data compression. Results are summarized below.

5.4.1 Hybrid Cosine/DPCM Compression

This technique consists of performing a simplified Fourier transform (cosine transform) on a row of 16 pixels, then characterizing the statistical distribution of the resulting coefficients in a column of 16 such rows (**figure 5-19**). The reason data compression works at all is that image data is not random row-to-row but is highly correlated. Data compression methods exploit this correlation.

The discrete cosine transform (DCT) acts as a scrambler or hash coder which decorrelates pixel data. The distribution of cosine series coefficients can then be compactly characterized in terms of a few statistics such as the mean and standard deviation.

The DPCM (differential pulse code modulation) process consists of predicting the DCT coefficients in 16 successive rows. The difference between predicted and actual values (error signal) is then transmitted using less data than required to send the actual coefficients. The combination of DCT and DPCM thus summarizes data in a window of 16-by-16 pixels in the image. The coherence of the visual world assures redundancy in such contiguous regions. Extended areas of color, texture, or boundary edges maintain trends across many pixels.

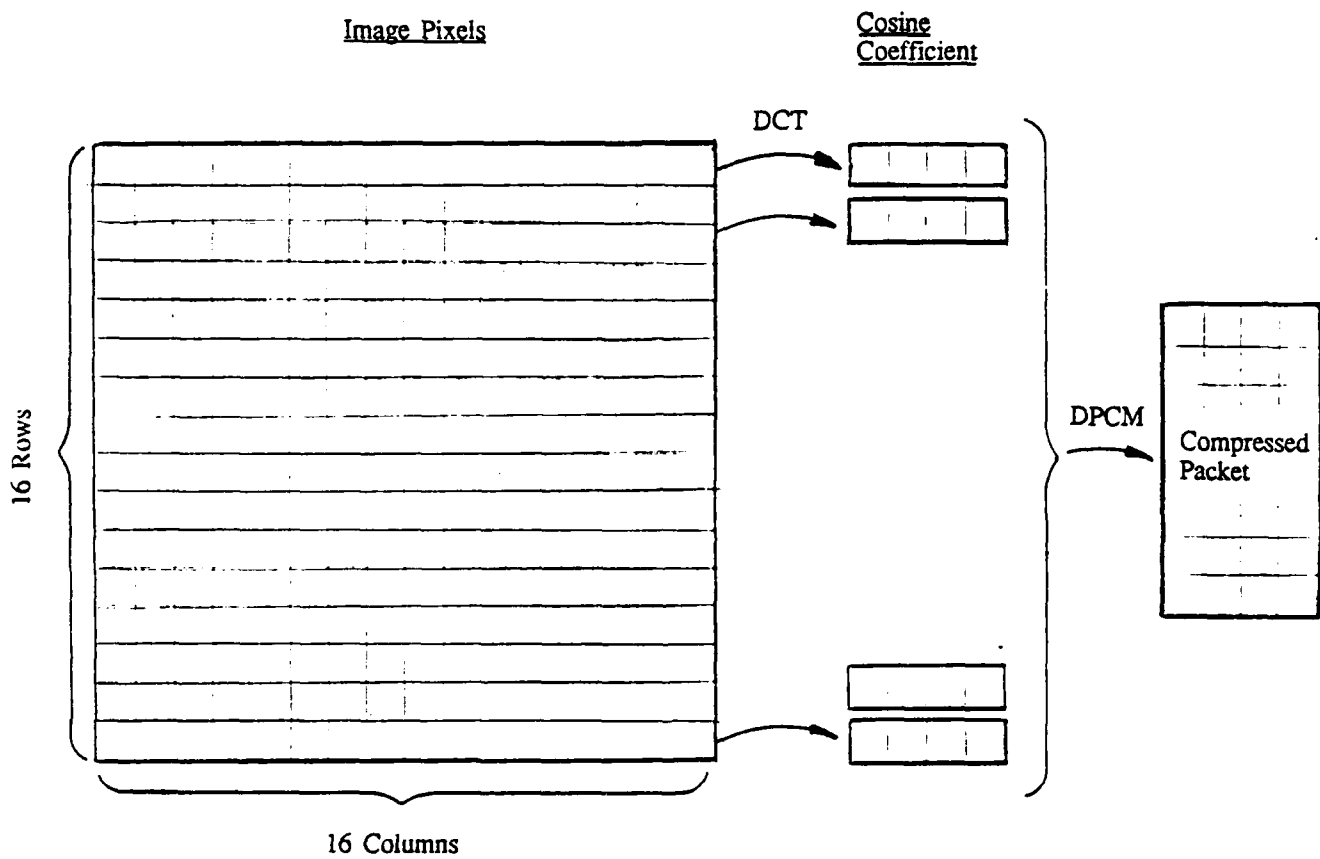


Figure 5-19. 16-by-16 Pixel Domain of Hybrid Data Compression

The net result of hybrid Cosine/DPCM coding is 8-to-1 compression of data. Reconstruction of original data from compressed data is accomplished by reversing each step of the hybrid process in reverse order. The fixed block length of this particular hybrid technique has a number of benefits for computation and transmission. Input computation can be carried out at realtime pixel rates using fixed bit length digital logic units. Processing delay is only 16 video lines required to construct a block. There is just enough data in the 16-by-16 window to exploit image redundancy. No global image content is required for computation. This method applies to grayscale, color components, or binary edge data. Errors in transmission are restricted to a single block, and are generally not correlated frame-to-frame, and will therefore cause very short image disruptions in small areas of the display.

5.4.2 Exotic High Compression Methods

In general, the higher the compression ratio, the higher the amount of processing needed to implement the technique. Intuitively, higher compression ratio techniques exploit redundancy over larger spans of data. Data interdependencies are therefore more complex, and encoding and decoding them requires more data comparison, pooling, and sorting. All of these require more computing power, which translates to long delays on small machines or large special purpose machines. Either solution reduces feasibility for mobile systems.

Extremely high compression ratio techniques yield data structures which are dependent on image content, and therefore variable length and variable structure. An example is quadtree coding. This variation on image pyramiding [Kawaguchi 80, 83] represents large quadrants of images with a single value if there is little variation within the quadrants. Thus an austere image can be represented with sparse data. Complex images, on the other hand, are represented with complex tree data structures. These have two drawbacks from the point of view of image transmission. The first is that reconstruction requires backtracking the tree, which can consume times which exceed video rates, or equipment too large to be portable. The second is that small errors in transmission induced by jamming or weak signals at extreme range, can have catastrophic effects. A minor transpose in tree data structure can scramble the whole image. Fixed length coding techniques require less processing and are more robust in the presence of noise.

The highest known compression ratios have been demonstrated in fractal data compression techniques described in [Barnsley 88]. These ingenious techniques represent images as growth algorithms, much like frost patterns on windows in winter. Complex natural scenes are described in terms of "seeds" and sequences of geometric transformations which are statistically developed. While these techniques may be useful in archiving massive image data libraries, processing requirements far exceed the constraints of realtime video transmission. Thousands of hours of VAX computer time are cited in compressing images at 10,000:1 ratios.

The size and processing delay of encoding and decoding equipment for most very high compression techniques are prohibitive for remote driving applications.

5.5 Summary

This study has proved the feasibility of a video image compression system which can provide full color, wide field of view, real time imagery with high central resolution. Compression ratios of 1600-to-1 are estimated, bringing transmission bandwidth down sufficiently to abolish line-of-sight restrictions.

The compression process is composed of three successive stages as follows:

- * Matching the display to human perceptual resolution provides better than 25-to-1 compression rate with no loss of perceived cues.
- * Separating perceptual channels into low resolution, high discrimination level color and high resolution, low discrimination level contrast edges yields 8-to-1 compression ratio.
- * Robust standard data compression techniques yield an additional 8-to-1 compression ratio.

Implementation is compact and robust and can be constructed on roughly 10 printed circuit boards approximately as complex as a digital video frame grabber.

A Phase II project is proposed for building and testing a realtime laboratory prototype for the image compression system, and for specifying a mobile field system based on the results of prototype testing and experimentation.

BIBLIOGRAPHY

- [Barnsley 88] Michael F., and Sloan, Alan D., "A Better Way to Compress Images", BYTE, January 1988, pp. 215-223.
- [Burt 83] Peter F., and Adelson, E. A., "The LaPlacian Pyramid as a Compact Image Code", IEEE Trans. on Communications, 31, April, 1983, pp. 532-540.
- [Business Week, 89] Cover story, "Super Television - High Definition TV", pp. 54-66, January 30, 1989.
- [Conrac 80] Raster Graphics Handbook, Conrac Division, Conrac Corp, 600 North Rimsdale Avenue,, Covina, Calif. 1980.
- [Faugeras 76] O. D., Digital Color Image Processing and Psychophysics Within the Framework of a Human Visual Model, Tech Report UTEC-CSc-77-029 (Ph. D. Thesis) DARPA Contract DAHC15-73-C-0363, Computer Science Dept, University of Utah, June 1976. Ref pp. 56.
- [Fisher 88] T, R. D. Juday, "Real-time video image remapper", Proceedings of the SPIE Conference on Pattern Recognition and Signal Processing, Vol 938, Digital and Optical Shape Representation and Pattern Recognition, Orlando, (1988).
- [Griswold 82] Norman and Sayoud, K., "Unsupervised Learning Approach to Differential Pulse Code Modulation", IEEE Trans. PAMI, July, 1982, 380-391.
- [Jung 73] R., "Visual Perception and Neurophysiology", in Handbook of Sensory Physiology, Vol VII/3A, Springer-Verlag, 1973.
- [Kawaguchi 80] Eiji and Tsutomu, E., "On a method of binary-picture representation and its application to data compression", IEEE Trans. PAMI 2, No. 1, Jan 1980.
- [Kawaguchi 80] Eiji, Tsutomu, E. and Matsunaga, J-I., "Depth-first picture expression viewed from digital picture processing" IEEE Trans. PAMI 5, No. 4, July, 1983.
- [Kraft 80] C.L., Anderson, C.D., and Elworth, C.L., Psychophysical Criteria for Visual Simulation Systems, Boeing Aerospace Final Report number AFHRL-TR-79-30 for USAF Human Resources Lab, AD A084 776, May 1980.

- [Livingston 88] Margaret and Hubel, David, "Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception", Science, Vol 240, 6-MAY-88, pp. 740-749.
- [Porat 88] M. and Zeevi, Y.Y., "The Generalized Gabor Scheme of Image Representation in Biological and Machine Vision", IEEE Trans. PAMI, Vol 10, No. 4, pp. 452-468, July 1988.
- [Schwartz 85] E. L., "On the Mathematical structure of the Retinotopic Mapping of Primate Striate Cortex", Science, Vol 277, pp 1066, 1985.
- [Weiman 79] C.F.R., and Chaikin, G., "Logarithmic Spiral Grids for Image Processing and Display", Computer Graphics and Image Processing, Vol. 11 Pages 197-226, 1979.
- [Weiman 87] C.F.R., "Tracking system applications of exponential sensor array system", Final report to NASA (JSC) on NASA Phase I SBIR 86-1, Contract number NAS 9-17728 (1987).
- [Weiman 88a], "Exponential sensor array geometry and simulation", Proceedings of the SPIE Conf. on Pattern Recognition and Signal Processing, Vol 938, Digital and Optical Shape Representation and Pattern Recognition, Orlando, April 88.
- [Weiman 88b], "3-D sensing with polar exponential sensor arrays", Proceedings of the SPIE Conf. on Pattern Recognition and Signal Processing, Vol 938, Digital and Optical Shape Representation and Pattern Recognition, Orlando, (1988).
- [Wilson 83] S. W. "On the retino-cortical mapping", Int'l Journal on Man-Machine Studies, (18) pp. 361-389, (1983).
- [Wilson 85] S.W., "Adaptive 'Cortical' Pattern Recognition", in Proceedings of International Conference on Genetic Algorithms and Their Applications, Carnegie-Mellon Univ., Pittsburgh, July 24-26, 1985.
- [Zeevi 88] Y.Y. and Daugman, J.G., "Some Psychophysical Aspects of Visual Processing of Displayed Information", in 1981 Image Generation/Display Conference II, USAF Human Resources Lab, pp. 260-277, June 10-12, 1988.

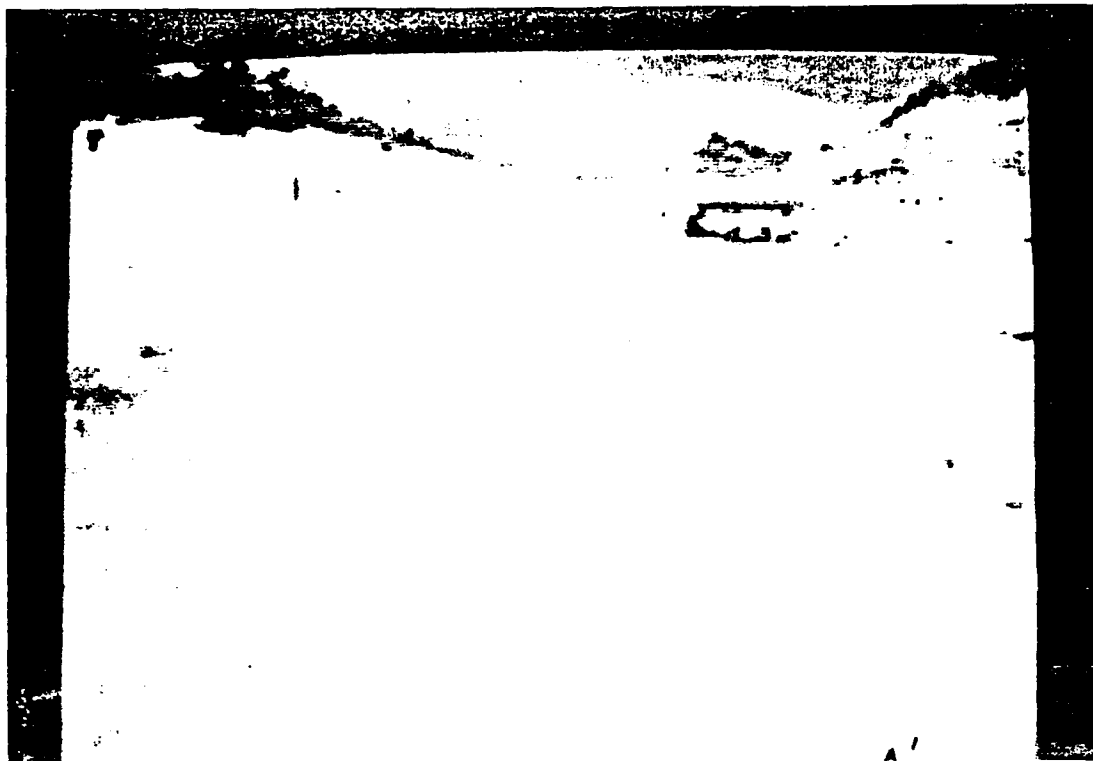


Photo 1 - Raw Video Color Image of Typical Driving Scene

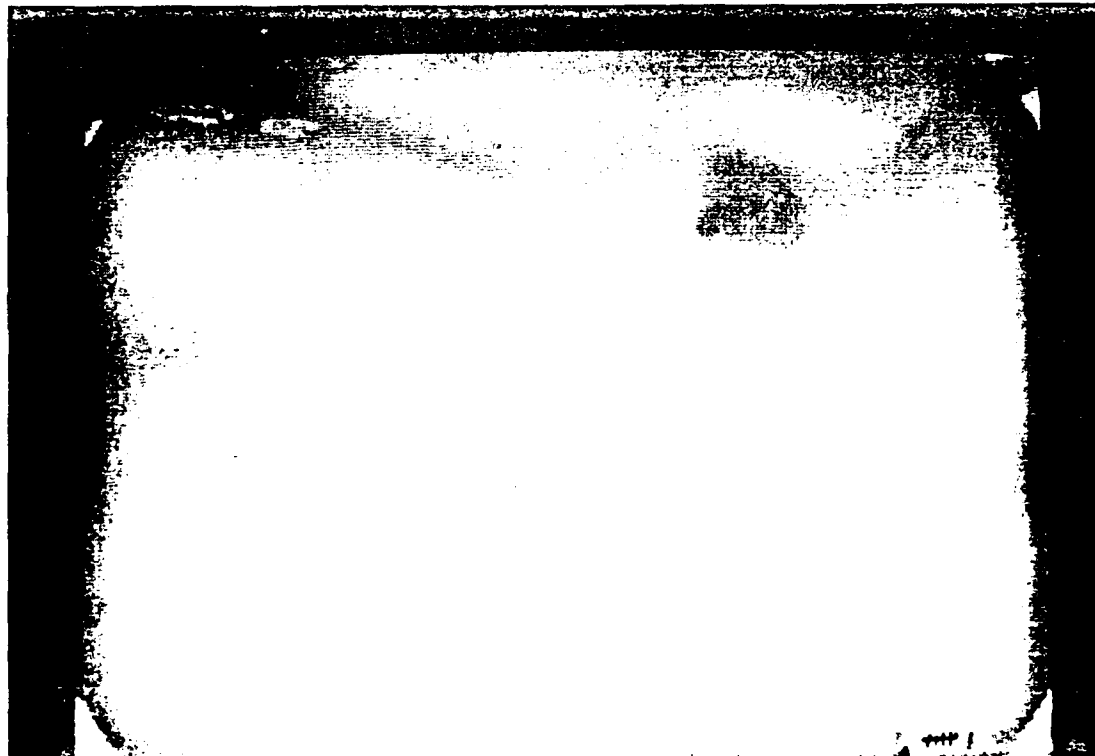


Photo 2 - Reconstructed Low Resolution Color Image



Photo 3 - Reconstructed Edge Density Image



Photo 4 - Composite Color and Edge Image



Photo 5 - Composite Image with Different Center of Attention



Photo 6 - Log-Mapped Image in Upper Left of Picture



Photo 7 - Reconstruction with Zeroth Order Interpolation



Photo 8 - Reconstructed High Resolution Color Image



Photo 9 - Reconstructed Edges Smeared into Blobs



Photo 10 - Gradient Image of the Smeared Edges



Photo 11 - Composite Image using Linear Feature Data

DISTRIBUTION LIST

	Copies
Commander Defense Technical Information Center ATTN: DDAC Bldg. 5, Cameron Station Alexandria, VA 22304-9990	12
Commander U.S. Army Tank-Automotive Command ATTN: AMSTA-DDL (Technical Library) Warren, MI 48397-5000	2
Commander U.S. Army Communications-Electronics Command ATTN: AMSEL-RD-ASCO-SC (Kobylarz) Fort Monmouth, NJ 07703	1
U.S. Army Armor Center ATTN: ATSB-CD-SD (CPT Szydlowski) Fort Knox, KY 40121	1
Commander U.S. Army Human Engineering Laboratory ATTN: SLCHE-CS (Holly) Aberdeen Proving Grounds, MD 21005-5001	1
Naval Ocean Systems Center ATTN: Mr. Tom Hughes Code 531 Kailua, HA 96734-0997	1
Martin Marietta Aero & Naval Systems ATTN: Scott Myers 103 Chesapeake Park Plaza Baltimore, MD 21220	1
FMC Corporation Central Engineering Laboratories ATTN: Lou McTamanev 1205 Coleman Avenue Box 580 Santa Clara, CA 95052	1
General Dynamics Land Systems Division ATTN: Bill Copen 6674 Crescent Green West Bloomfield, MI 48333	1

DISTRIBUTION LIST

	Copies
General Dynamics Land Systems Division ATTN: Detlef Gersdorff P.O. Box 2074 Warren, MI 48090-2074	1
KFO Associates, Inc. ATTN: Jerry Lewis 6 Pearl Court Allendale, NJ 07401	1
Texas A&M University Electrical Engineering Department ATTN: Norman Griswold College Station, TX 77840	1